

```

US-09-985-689A-3 (1-433) x TA315H10P (1-574)
QY 18 GlyLeuThrGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 63 GGTATGACGGTAGTGGTGGAGATATAGGGGTAGCGATACGGGTATCGACTTT-----116
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuThrAlaLeuGly 57
Db 117 AACAGCTGTTTCTTCCAGATCCA-----AATCAAGAGGTGGCGCTTTATCCCGAAGTT 170
QY 58 ArgThrAsnAsn-----61
Db 171 AACTATACACCGCAAAATCGTGTCAATTCGCCCGTGTGACTTCATCCGGGGGATATAC 230
QY 62 -----AlaAsnAspProAsnGlyHisGlyThrHisVal-----AlaGlySer 75
Db 231 TTTGTGGGGATGAGAGAAATAGGTGATGCGACGACGTGGCAGGTACCGCACGGGGAGT 290
QY 76 ValLeu-----GlyAsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 291 GTATTATGTAACACGGTAAATGCCAAGTATAATGTGTGCGCAAGGGGGCGNAGATTTC 350
QY 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThr 112
Db 351 TTCAGGGGTGTGGGTGCCCATCCAGTCAGAGCTTGTCTCCCGCCACGAGCTTACTCAA 410
QY 113 LeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaPro 132
Db 411 ATTATTCGTCGGGATATGGCGTGGAGCCCGTGTGTCTCAAACTCGGGGTTTGT 470
QY 133 ValAsnGlyAlaThrAlaAsnSerArgGlnValAspGluThrValArgAsn---Asn 151
Db 471 GTCCTCCCGAGTATTCTGTGGAAAGGATATGGATGAGTTTGGCAGTATGATGAC 530
QY 152 AspMetThrValLeuPheAlaAlaGlyAsn 161
Db 531 GATGGCTACTTATCTTCCACTGGCAAC 560

RESULT 12
AJ273402 614 bp mRNA linear EST 29-DEC-1999
LOCUS AJ273402 Metarhizium anisopliae ARSEF 2575 Metarhizium anisopliae
DEFINITION cDNA clone Ma#948, mRNA sequence.
ACCESSION AJ273402.1 GI:6432774
VERSION AJ273402.1
KEYWORDS EST.
SOURCE Metarhizium anisopliae
ORGANISM Metarhizium anisopliae
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Clavicipitaceae; Clavicipitaceae;
Clavicipitaceae; Metarhizium.
REFERENCE 1 (bases 1 to 614)
AUTHORS Screen,S.E., Mathur,P. and St. Leger,R.J.
TITLE EST analysis of the insect pathogenic fungus Metarhizium anisopliae
JOURNAL Unpublished (1999)
COMMENT Contact: Screen SE
Entomology
University of Maryland
4112 Plant Sciences Building, College Park, MD 20742, USA.
FEATURES
Source 1..614
/organism="Metarhizium anisopliae"
/mol_type="mRNA"
/strain="ARSEF 2575"
/db_xref="taxon:5530"
/clone="Ma#948"
/note="Vector: UniZap; Metarhizium anisopliae was grown on insect cuticle for 24 hours. A cDNA library was constructed in the unidirectional Lambda vector, UniZap"
ORIGIN
Alignment Scores:

```

```

Pred. No.: 1,09e-08 Length: 614
Score: 188.00 Matches: 68
Best Similarity: 45.68% Conservative: 43
Best Local Similarity: 27.98% Mismatches: 84
Query Match: 8.31% Indels: 48
Db: 9 Gaps: 12

US-09-985-689A-3 (1-433) x AJ273402 (1-614)
QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 11 GGTGAGGTACTTGCCTATATATATATATATATATATATATATATATATATATATATAG 52
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuThrAlaLeuGlyArgThrAsn 60
Db 53 GCTCCACACCCAGTTTGGGTGCGGCCACTTTCTTAAGAGTTCATCAGCGTCAAA 112
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 113 AAC---ACTGATGCCACGCCCATGGGACTCACTGGCTGGTACCATT-----GGTAGC 163
QY 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer--- 99
Db 164 AGACCTACCGTGTTCGCAAAAGGTAACTATGTTGTCAGGTTCTTGACACACAG 223
QY 100 ---SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrp 118
Db 224 GGCAGTGTCTTACTACTCCGGTATCATCATGATGGATGCTTGTGTCAGGACTCCAAAG 283
QY 119 AsnAlaGly-----AlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGly 135
Db 284 ACCGCGCGCTGCCCAACGGCGCCATTGCTTCCATGATGAGCTGGGA-----GGT 331
QY 136 AlaTyrThrAlaAsnSerArgGlnValAspGluThrValArgAsnAsnAspMetThrVal 155
Db 332 GGCTACTCGCGCTCCCAACCAAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 391
QY 156 LeuPheAlaAlaGlyAsnGlyProAsnSerGlyThrIleSerAlaProGlyThrAla 175
Db 392 GCGTCTCGCGCTGCCCAACCAAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 448
QY 176 LysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAla 195
Db 449 CCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 478
QY 196 AspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIle 215
Db 479 GAAATGACAGCGCATCTTCTTCTTCACTAC-----GGCAGAGTT 520
QY 216 LysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPro 235
Db 521 ---GTGATATTTTCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 559
QY 236 AspSerSerPheThrAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMet 255
Db 560 -----TGG-----ATTGGTGGCGGCAACCAACCATCTCTGTGCTGCTGCTGCTGCT 601
QY 256 AlaThrPro 258
Db 602 GCTACTCTCC 610

RESULT 13
BZ424995/c 2121 bp DNA linear GSS 13-DEC-2002
LOCUS BZ424995 Aspergillus terreus random genomic DNA clone library
DEFINITION 10023066-5726 Aspergillus terreus genomic, genomic survey sequence.
ACCESSION BZ424995.1 GI:26666450
VERSION BZ424995.1
KEYWORDS GSS.
SOURCE Aspergillus terreus
ORGANISM Aspergillus terreus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

```

REFERENCE 1 (bases 1 to 2121)
 AUTHORS Askenazi,M., Driggers,E.M., Holtzman,D.A., Norman,T.C., Iverson,S., Zimmer,D.P., Boers,M-E., Blomquist,P.R., Martinez,E.J., Monreal,A.W., Feibelman,T.P., Mayorga,M.E., Maxon,M.E., Sykes,K., Tobin,J., Cordero,E., Salama,S.R., Trueheart,J., Royer,J.C. and Madden,K.T.
 TITLE Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing strains
 JOURNAL Unpublished (2002)
 COMMENT Contact: Zimmer DP
 Microbia, Inc.
 One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
 Tel: 617-621-8322
 Fax: 617-621-8322
 Email: dzimmer@microbia.com
 Class: plasmid ends.
 Location/Qualifiers
 1. 2121
 /organism="Aspergillus terreus"
 /mol_type="genomic DNA"
 /strain="ATCC 20542 (A. terreus Thom, anamorph)"
 /db_xref="taxon:33178"
 /lab_host="Escherichia coli"
 /clone_lib="Aspergillus terreus random genomic DNA clone library"
 /note="vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.66e-07 Length: 2121
 Score: 185.50 Matches: 79
 Percent Similarity: 38.97% Conservative: 34
 Best Local Similarity: 27.24% Mismatches: 86
 Query Match: 8.20% Indels: 92
 DB: 28 Gaps: 12

US-09-985-689a-3 (1-433) x B2424995 (1-2121)

QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 1281 GCGAGGACACTAGCTAGCTGTAGACACCGGATCACGTGGAC-----1234

QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsn 60
 DB 1233 -----CACGAGGAATTTAGCGCGCTCGAGTCTCGGTACACCGCTCGCGTGGCAG 1180

QY 61 AsnAlaAsnApproAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
 DB 1179 CATGTG---GACGGGTGCGGCATGCGCCACCGTTCGTGTACAAATGGTGGTAACA 1123

QY 81 LeuAsnLysGlyMetAlaProGlnAlaAlaAsnLeuValPheGlnSerIle-MetAspSerSe 100
 DB 1122 TAC-----GGGTGGCCCAAGAGCCCAACTGCTG-----TCGGTCAAGGTCTTCGTC 1075

QY 100 rGlyGlyLeuGlyLeuProSerAsnLeuAsn-----ThrLe 113
 DB 1074 GGGGAATCGAGCAGCACATCCATCATCTCTGGACGGCTTCAACCTGGGTATGTTATTCTGC 1015

QY 113 uPheSerGln-----AlaTrpAsnAlaGlyAlaAr 123
 DB 1014 GTTTCACAGACGGCACTGGAAAGCGCAGGATCAACATGAGCTTGGGTATGTTATTCTGC 955

QY 123 gileHisThrAsnSerTrp-----GlyAlaProValAsnGlyAlaTy 137
 DB 954 CATCTATGTGATGA-TATCATTTGGAGTCACACAGCTGACGCAAGCATAGCGGTGGATA 896

QY 137 rThrAlaAsnSerArgGlnValAspGlyTyValArgAsnAsnAspMetThrValLeuPh 157
 DB 895 CTCCAAGGCCTCAATGATGCGGTGAGACCGCTTCAACGAGGAGGCTCTCCATCGT 836

QY 157 eAlaAlaGlyAsnGluGly-----ProAsnSerGlyThrIleSer-- 170

Db 835 CCCCCCGGCAATGAGATGTGGTGCCACGCTTCACCGCTGTGATGGAGTGGCGACTCC 776
 QY 171 -----AlaProGly-ThrAlaLysAlaIai 179
 Db 775 GGTGACTGATTCGACAGACGCGCTCGAGCAGCGCGGCTTCTGCTCTGATGCC 716
 QY 179 leThrValGlyAla-----ThrGluAsnTyArgProSerPheGlySerI 194
 Db 715 TTACCGTGGTGGCATCAACGTGAACAACACCGCTGCTATTTCTCCAACTACGGCTCG 656
 QY 194 leAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyA 214
 Db 655 TGGTG-----651
 QY 214 rGlyLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuA 234
 Db 650 -----GATATCTTCGCGCGCGGCGGCGAGACATCTCTCTGCC-----615
 QY 234 laProAspSerSerPheTrpAlaAsnTyArgAsnSerLysTyAlaTyMetGlyGlyThrS 254
 Db 614 -----TGGATCGGCTCCAAACACGCGCCACACCATCTCGGCGACCT 572
 QY 254 erMetAlaThrProIleValAlaGly 262
 Db 571 CCATGCGCCACCCCGCCACATTGTGCGC 546

RESULT 14
 BQ770462 665 bp mRNA linear EST 26-JUL-2002
 UI-M-F10-byv-m-19-0-UI-r1 NIH_BMAP_F10 Mus musculus cDNA clone
 IMAGE:5702970 5', mRNA sequence.
 BQ770462
 BQ770462.1 GI:21978936
 EST.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC http://mgi.mcg.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-r@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pYX-5.
 Location/Qualifiers
 1. 665
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:5702970"
 /tissue_type="whole brain"
 /dev_stage="embryo 12.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH BMAP_F10"
 /note="Organ: Brain; Vector: pYX-Asc; Site1: EcoR I;
 Site 2: Not I; The library was constructed according
 Bernaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated

FEATURES
 source

with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCACGAC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

Alignment Scores:
Pred. No.: 4,59e-08 Length: 665
Score: 182.50 Matches: 65
Percent Similarity: 42.68% Conservative: 37
Best Local Similarity: 27.20% Mismatches: 96
Query Match: 8.06% Indels: 41
DB: 13 Gaps: 7

US-09-985-689A-3 (1-433) x BQ770462 (1-665)

```

QY 63 AsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsn 82
DB 14 GATGATGGCTAGGCGATGGACATTCGTGAGGTGTGATGCCAGCATGAGGAGTGC 73
QY 83 LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSerGly 101
DB 74 CAAGGATTTGCTCCAGATGCAGAGCTGCACATCTTCAGGGTCTTTACCAACAATCAG--- 130
QY 102 GlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaThrPheAsnAlaGly 121
DB 131 -----GTGCTTACACATCTTCGTTTCTGGATGCTTCAACTATGCC 172
QY 122 AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaThrAlaAsnSer 141
DB 173 ATCTTAAGAAGATGACGCTTCTCAACCTAGCATCGTGGCCCGCATTCATGATCAT 232
QY 142 ArgGlnValAspGluTyr-----ValArgAsnAsnAspMetThrValLeuPheAla 159
DB 233 CGGTGTTGTTGACAAGGTGGGAATTAACAGCATCAATGTAATATGTTCTCTATT 292
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLeuAlaIle 179
DB 293 GGCATATGATGGACCTCTCTATGGCACTCTGATATACCTGCTGATCAGATGATGATT 352
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
DB 353 CGAGTGGGTGGC-----ATTGACTTTGAAGAT 379
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArg-----AspGly 213
DB 380 AACATCGCTCGCTTTCTTCCAGGGGAATGACTACTCTGGGAATACCAGGAGCTATGTT 439
QY 214 ArgIleLeuProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerLeu 233
DB 440 CGTGTGAAGCTGCACATTCACCTATGCTGCTGGAGTGGGGGT----- 484
QY 234 AlaProAspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaThrMetGlyGlyThr 253
DB 485 -----TCCGGTGTGAAGGGGGCTGGCTGCACCTCTCAGGGACC 523
QY 254 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheLeuLys 273
DB 524 AGTGTGCTTCNCCAGTGGTGGTGGGGCGTGCACCTGTTAGTAGCAGCAGTACAGAAG 583
QY 274 AsnArgGlyThrProLysProSerLeuIleLysAlaAlaLeuIleAlaGlyAla 292
DB 584 CGGAGAGCTGGTG-----AATCTGCCAGTGTGAAGCAGCTTTGTAGCGTCAGCC 634

```

RESULT 15
BZ8933395/c 716 bp DNA linear GSS 30-JUL-2003
LOCUS HL2_0177 H1 pUC18 Library Halorubrum lacusprofundi genomic 5',
DEFINITION genomic survey sequence.

ACCESSION BZ8933395
VERSION BZ8933395.1 GI:33343985
KEYWORDS GSS.
SOURCE Halorubrum lacusprofundi
ORGANISM Halorubrum lacusprofundi
Archaea: Euryarchaeota; Halobacteria; Halobacteriales;
Halobacteriaceae; Halorubrum.

REFERENCE 1 (bases 1 to 716)
AUTHORS Goo Y., Reach, J., Glusman, G., Baliga, N.S., Deutsch, K., Pan, M.,
Dassarma, S., Ng, W.V. and Hood, L.
TITLE Low-pass Sequencing for Microbial Comparative Genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Goo Y
Institute for Systems Biology
1441 North 34th Street, Seattle, WA 98103, USA
Tel: 206 732 1412
Fax: 206 732 1299
Email: ygoos@systemsbiology.org
Seq primer: M13 Forward
Class: shotgun.

FEATURES
Location/Qualifiers
1..716
/organism="Halorubrum lacusprofundi"
/mol_type="genomic DNA"
/strain="ATCC 49239"
/db_xref="taxon:2247"
/clone_lib="H1 pUC18 Library"
/notes="Vector: pUC18; Site_1: SmaI; A shotgun library was
constructed from Halorubrum lacusprofundi genomic DNA
using pUC18/SmaI/BAP plasmid"

ORIGIN

Alignment Scores:
Pred. No.: 5,21e-08 Length: 716
Score: 182.50 Matches: 71
Percent Similarity: 45.45% Conservative: 34
Best Local Similarity: 30.74% Mismatches: 66
Query Match: 8.06% Indels: 61
DB: 28 Gaps: 11

US-09-985-689A-3 (1-433) x BZ8933395 (1-716)

```

QY 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu---Gly 78
DB 701 TCCGATGCTCGGACGACGAGCGGCGGAACTCACGTCCGGGAACCGTCCGGCGGA 642
QY 79 AsnAlaLeuAsnLys-----GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 96
DB 641 AACGCGAGTGGGACCGCAATCGCGCTCGCACCGACGCGTCCCGCGTCCGCGCG 582
QY 97 MetAsp-SerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerG 116
DB 581 TTCGACGACGACGGA-----CCACGCGACGCTCGTCCCGCTCGTCCGCG 537
QY 116 nAlaTrpAsnAlaGlyAlaArgIleHisThrAsnSerTrp-----GlyAlaProva 133
DB 536 GGAATGGAACACGCGACGAGGATCCGACGCTGCTTCAGATGAGCCCTCGGTGCG 477
QY 133 LAsnGlyAlaThrThrAlaAsnSerArgGlnValAspGluTyrValArgAsnAsnAspMe 153
DB 476 GACGCGGACCTTACATCTACTTCATCGAACCG-GTT---CGGAACACTCGCAGTCCGGGAA 421
QY 153 tThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProG 173
DB 420 GATCGCGTCTGTTTCGGCGGGAACATCGGT-----CAAGGAACGTCGAGTTCCTCCCG 367
QY 173 yThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySe 193
DB 366 GAACGCTTACGACTCGCTCGGCTCGGAGCGGTCAACGACGCGC----- 321
QY 193 rIleAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThr----- 210
DB 320 -----GGCGTCCGCGGACTTCTCCAGC---GGAGAGACGATCAACAC 283

```

```

Qy 211 -----ArgAspGlyArgIleLeuProAs 218
Db 282 GTCAGCGCGTGGGAAAGTATGCCCGCGGACTGGCCCGACGAGTACGTGGTTCGGA 223
Qy 218 pValThrAlaProGlyThrPheIleLeuSerAlaArg-----SerSerLeuAlaPr 235
Db 222 CGTGTGCGCGCCAGCGCGTACGTATCTCGGCGGAACCGGCGGTACAAACATCCGAAA 163
Qy 235 oAspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMe 255
Db 162 GGAC-----GGCACCTCCAT 148
Qy 255 tAlaThrProIleValAlaGlyAsnValAla 265
Db 147 GGCCGACCGCACGTCAGCGCGCTCGCGGCG 117

```

Search completed: April 4, 2004, 11:54:47
Job time : 2284.58 secs


```

Qy 119 AsnAlaGly-----AlaArgileHisThrAsnSerTrpGlyAlaProValAsnGly 135
Db 284 ACCCGGGCTGCCACACGGCGCATTCCTCCATGAGCCTGGGA-----GGT 331

Qy 136 AlaTyrThrAlaAsnSerArgGlnValAspGluTyrValArgAsnAspMetThrVal 155
Db 332 GGCTACTCGCGCTCGCAACCAAGGCTGCTGTGTGCTCAATCTCGTGTCTTCCTT 391

Qy 156 LeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAla 175
Db 392 GCCGTCGCCGCTGCCACAGATACCGCGGATGCCAGACACCTCT---CCCGTTCGGAG 448

Qy 176 LysAlaAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAla 195
Db 449 CCTTCGCCCTGCATGTGTGCTCT---GCG 478

Qy 196 AspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgile 215
Db 479 GAAATGACAGCGCATCTTCCTTCTCACTAC-----GGCAGAGTT 520

Qy 216 LysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPro 235
Db 521 ---GTCGATATTTTCGCTCTCTGTAGCAATGTTCTTCCACC-----559

Qy 236 AspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMet 255
Db 560 -----TGG-----ATTGGTGGCGGCACAAACACCATCTCTGTACTCTNATG 601

Qy 256 AlaThrPro 258
Db 602 GCTACTCCC 610

RESULT 13
TA315H10P 574 bp DNA linear GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 315h10, forward sequence,
DEFINITION genomic survey sequence.
ACCESSION AL490202
VERSION AL490202.1 GI:11866292
KEYWORDS GSS.
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei
REFERENCE 1 (bases 1 to 574)
AUTHORS Hall N., Bowman S., Lennard N.J., Doggett J., Atkin R.,
Chillingworth C., Ormond D., Harris B., El-Sayed N., Hou L.,
Melville S.E., Rajandream M.A. and Barrell B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: neilsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES
Location/Qualifiers
1..574
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="315h10"
ORIGIN

```

Alignment Scores:

```

Pred. No.: 9.18e-09 Length: 574
Score: 188.00 Matches: 53
Percent Similarity: 47.65% Conservative: 28
Best Local Similarity: 31.18% Mismatches: 59
Query Match: 8.32% Indels: 30
DB: 29 Gaps: 7

```

US-09-985-689A-4 (1-433) x TA315H10P (1-574)

```

Qy 18 GlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 63 GGTATTGACGGTAGTGGTAGATATATAGGGTAGCCGATACGGTATCGATTT-----116

Qy 38 AsnAspSerSerMetHisLualaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 117 AACAGCTGTTTCTTCCAGCATCCA-----AATCAAGAGGTGGCGCTTTACCCGAAGTT 170

Qy 58 ArgThrAsnAsn-----61

Db 171 AACTATAACCCCGCAAAATCGTGTCAATTCGCCCGTGTGACTTCATCCGGGGGATTAC 230

Qy 62 ---AlaSerAspProAsn---GlyHisGlyThrHisVal-----AlaGlySer 75
Db 231 TTTGCTGGGGATGAGGAATAGGTTCATGCGACGCGCAGGTGCGAGGTACCGCAGCGGGAGT 290

Qy 76 ValLeu-----GlyAsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 291 GTTATTAGTAACGCGGTAAATGCCAAGTATAATGTTGCGCAAGGGGCGGAGATTTC 350

Qy 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThr 112
Db 351 TTCAGGGGGTGGCGTCCCATCCAGTCAGAGGTGTCCTCCCTCCACAGCTTACTCAA 410

Qy 113 LeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro 132
Db 411 ATTATTCTCCCGGATATGCGCTGAGCCCGCTGTGTTCTCAAACTCGTGGGGTTTGT 470

Qy 133 ValAsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGluTyrValArgAsn---Asn 151
Db 471 GTCCTCTCGAGTATTCGCTGTGGAAAAGGATATGGATGGATGAGTTGCGAGTAGTTATGAC 530

Qy 152 AspMetThrValLeuPheAlaAlaGlyAsn 161
Db 531 GATCGCTACTTATCTCTCCACTGGCAAC 560

```

RESULT 14

BZ893395/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BZ893395 716 bp DNA linear GSS 30-JUL-2003
 HL2_0177 H1 pUC18 Library Halorubrum lacusprofundi genomic 5',
 genomic survey sequence.

BZ893395

GSS.

Halorubrum lacusprofundi

Halorubrum lacusprofundi

Archaea; Euryarchaeota; Halobacteria; Halobacteriales;

Halobacteriaceae; Halorubrum.

1 (bases 1 to 716)

Goo Y., Roach J., Glusman, G., Baliga, N.S., Deutsch, K., Pan, M.,

Dassarma, S., Ng, W. V. and Hood, L.

Low-pass Sequencing for Microbial Comparative Genomics

Unpublished (2003)

Contact: Goo Y

Institute for Systems Biology

1441 North 34th Street, Seattle, WA 98103, USA

Tel: 206 732 1412

Fax: 206 732 1299

Email: ygoo@systemsbiology.org

Seq primer: M13 Forward

Class: shotgun.

Location/Qualifiers

```
source
1..716
/organism="Halorubrum lacusprofundi"
/mol_type="genomic DNA"
/strain="ATCC 49239"
/db_xref="taxon:2247"
/clone_lib="HL pUC18 Library"
/note="Vector: pUC18; Site 1: SmaI; A shotgun library was constructed from Halorubrum lacusprofundi genomic DNA using pUC18/SmaI/BAP plasmid"
ORIGIN
Alignment Scores:
Pred. No.: 2,428-08 Length: 716
Score: 185.50 Matches: 72
Percent Similarity: 45.45% Conservative: 33
Best Local Similarity: 31.17% Mismatches: 66
Query Match: 8.21% Indels: 61
DB: 28 Gaps: 11
US-09-985-689A-4 (1-433) x BZ893395 (1-716)
QY 60 AsnAlaSerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu---Gly 78
Db 701 TCCGATGCGTCGACGACGCGGACGGAACCTACGTCGCGGAAACCGTCGCGGCGGA 642
QY 79 AsnAlaLeuAsnLys-----GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 96
Db 641 AACGGAGTGGACCGCAATCGCGCTCGACCGACGCGTGCCTCCACGGATCAAGGTG 582
QY 97 MetAsp-SerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerG 116
Db 581 TTCGACGACGACGGGA-----CCAAACGGGAGCGTTCGTCGCGGTCGTCGCG 537
QY 116 AlaThrProAlaGlyAlaArgIleHisThrAsnSerTrp-----GlyAlaProVa 133
Db 536 GGAATGGAACACCGGACGCGGACGTCGCGGACGTCGTCGATGAGCGCTCGTCGCG 477
QY 133 LAsnGlyAlaTyThrAlaAsnSerArgGlnValAspGluTyrValArgAsnAsnAspMe 153
Db 476 GACGGGCACTTACCACTTCATCGAACCG-GTT---CGGAACACTCGAGTCGCGGAA 421
QY 153 tThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProG 173
Db 420 GATCGCGGTGCTTTCGCGCGGGAACATCGGT-----CAAGGAACGTCGAGTCTCCCGG 367
QY 173 yThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyArgProSerPheGlyse 193
Db 366 GAACGCTACGACTCGCTCGCGTCGAGCGGTCAACGACACGCCG----- 321
QY 193 rIleAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThr----- 210
Db 320 -----GGCGTCGCGCACTTCTCCAGC---GGAGAGACGATCAACAC 283
QY 211 -----ArgAspGlyArgIleLysProAs 218
Db 282 GTCAAGCGCGTGGGAGTATGATCCCGCGGACTGCGCGGACGAGTACGTGTTCCGGA 223
QY 218 pValThrAlaProGlyThrPheIleLeuSerAlaArg-----SerSerLeuAlaPr 235
Db 222 CGTGTGCGCGCCAGGCGTCGAGCGTATACTCGCGGAAACCGCGGTCACAAACATCCGAA 163
QY 235 oAspSerSerPheTrpAlaAsnTyArgAsnSerLysTyAlaTyMetGlyGlyThrSerMe 255
Db 162 GGAC-----GGCACCTCCAT 148
QY 255 tAlaThrProIleValAlaGlyAsnValAla 265
Db 147 GGCGGACCGCACGTCAGCGCGCTCGCGGCG 117
RESULT 15
BQ770462 665 bp mRNA linear EST 26-JUL-2002
LOCUS BQ770462
DEFINITION UI-M-F10-byy-m-19-0-UI.r1 NIH_BMAP_F10 Mus musculus cDNA clone
```

```
IMAGE:5702970 5', mRNA sequence.
BQ770462
VERSION BQ770462.1 GI:21978936
EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 665)
NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
AUTHORS Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@bbsr@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1..665
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5702970"
/tissue_type="whole brain"
/dev_stage="embryo 12.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP_F10"
/note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CACCCACGAC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."
```

```
ORIGIN
Alignment Scores:
Pred. No.: 5,5e-08 Length: 665
Score: 181.50 Matches: 65
Percent Similarity: 42.44% Conservative: 36
Best Local Similarity: 27.31% Mismatches: 96
Query Match: 8.03% Indels: 41
DB: 13 Gaps: 7
US-09-985-689A-4 (1-433) x BQ770462 (1-665)
QY 64 AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsnLys 83
Db 17 GATGGGTAGGCCATGCGACATCGTTGAGGTGATGCCAGCATGAGGAGTGCCAA 76
QY 84 GlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSerGlyGly 102
Db 77 GGATTTCCTCCAGATCGAGAGCTGCACATCTTCAGGTCTTTTACCACCAATCAG----- 130
QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
Db 131 -----GTGCTTACACATCTGTGTTCTCGATGCTTCACTGATCCATC 175
```

```
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
Db 176 CTAAGAAGATGGACGTTCTCAACCTTAGCATGGTGGCCGCCACTTCATGGATCATCCG 235
Qy 143 GlnValAspGluTyr-----ValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
Db 236 TTTCCTTCACAAGGTGTGGGAATTAAACACCAACAAATGTAATTATGTTCTGCTATTGGC 295
Qy 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 296 AATGATGGACCTCTCTATGGCACTCTGAATAACCTCTGATCAGATGGATGTGATTGGA 355
Qy 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
Db 356 GTGGGTGGC-----ATTGACTTTGAAGATAAC 382
Qy 201 IleAlaGlnPheSerSerArgGlyAlaThrArg-----AspGlyArg 214
Db 383 ATCGCTCGCTTTCTTCAGGGGAATGACTACCTGGGAATTACCAGGAGGCTATGGTCGT 442
Qy 215 IleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAla 234
Db 443 GTGAAGCCTGACATGTACCTATGTTGCTGGAGTGGGGGT----- 484
Qy 235 ProAspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSer 254
Db 485 -----TCCGGTGTGAAGGGGGCTGCCGTGCACCTCTCAGGACCCAGT 526
Qy 255 MetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsn 274
Db 527 GTCGCTTCNCAGTGGTGGTGGGGCCGTACCTTGTAGTACACAGTACAGAACCGG 586
Qy 275 ArgGlyIleThrProLysProSerIleLysAlaAlaLeuIleAlaGlyAla 292
Db 587 GAGCTGGTG-----AATCCTGCCAGTGTGAAGCACGCTTTGTAGCGTCAGCC 634
```

Search completed: April 4, 2004, 11:54:55
Job time : 2282.58 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 16:47:19 ; Search time 3544.27 Seconds
(without alignments)
5295.174 Million cell updates/sec

Title: US-09-985-689A-5
Perfect score: 2261
Sequence: 1 NDVARGIVKADVAQNNYGLY.....EVQAYNVPSGPQPSLAIVH 433

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh

-Q=/cgr2_1/USPTO.spool/US09985689/runat_31032004_161806_4155/app_query.fasta_1.3498
-DB=GenEmbl -QFMT=fastpct -SURFIX=rge -MINMATCH=0.1 -LOOFCU=0 -ILOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 @CCN 1 1 11925 @runat_31032004_161806_4155 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl :
1: gb_ba :
2: gb_hgt :
3: gb_in :
4: gb_em :
5: gb_ov :
6: gb_pat :
7: gb_ph :
8: gb_pl :
9: gb_pr :
10: gb_ro :
11: gb_sts :
12: gb_sy :
13: gb_un :
14: gb_vi :
15: em_ba :
16: em_fun :
17: em_hum :
18: em_in :
19: em_mu :
20: em_om :
21: em_or :
22: em_ov :
23: em_pat :
24: em_ph :
25: em_pl :
26: em_ro :
27: em_sts :
28: em_un :

29: em_vi :
30: em_hgt_hum :
31: em_hgt_inv :
32: em_hgt_other :
33: em_hgt_mus :
34: em_hgt_pln :
35: em_hgt_rtd :
36: em_hgt_nam :
37: em_hgt_vrt :
38: em_sy :
39: em_hgt_hum :
40: em_hgt_mus :
41: em_hgt_other :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2261	100.0	1299	1	AB046405 Bacillus
2	2251	99.6	1299	1	AB046402 Bacillus
3	2245	99.3	2218	6	E03808 DNA encodin
4	2240	99.1	1299	1	AB046404 Bacillus
5	2044	90.4	3003	6	AR069954 Sequence
6	2044	90.4	3003	6	BD062155 Nucleic a
7	2024.5	89.5	1302	1	AB046406 Bacillus
8	2003.5	88.6	1923	1	AB084155 Bacillus
9	2002.5	88.6	1920	1	AB046403 Bacillus
10	2002.5	88.6	1920	6	AR368116 Sequence
11	1999.5	88.4	1923	6	AR368118 Sequence
12	1998.5	88.4	1305	6	AX839476 Sequence
13	1998.5	88.4	1923	1	AB051423 Bacillus
14	1998.5	88.4	1923	6	AR368117 Sequence
15	612.5	27.1	60006	1	AF268611 Unculture
16	558.5	24.7	301205	1	AE017218 Geobacter
17	501	22.2	5890	3	DDU60086 Dictyosteli
18	465	20.6	5772	3	AF466309 Dictyosteli
19	463.5	20.5	6115	3	DDU20432 Dictyosteli
20	443	19.6	1977	6	AR201152 Sequence
21	441.5	19.5	5109	3	AF263455 Dictyosteli
22	442.5	18.7	1236	6	AR201146 Sequence
23	422.5	18.7	1962	6	AR201155 Sequence
24	422.5	18.7	12452	1	AE010265 Pyrococcu
25	402.5	17.8	14376	1	AE013049 Thermoana
26	399	17.6	126928	2	AC096673 Trypanoso
27	395.5	17.5	303450	1	SC0939130 Streptomy
28	377.5	16.7	299850	1	AP004601 Oceanobac
29	370.5	15.5	2539	6	AR202322 Sequence
30	347.5	15.4	3348	1	D83672 Streptomyce
31	345.5	15.3	1239	1	AF305633 Thermoana
32	345.5	15.3	1239	1	AY028704 Thermoana
33	345.5	15.3	2809	6	AR202321 Sequence
34	345.5	15.3	3900	1	AB007809 Streptomy
35	344	15.2	292550	1	AP001513 Bacillus
36	343.5	15.2	12295	1	AE013026 Thermoana
37	341.5	15.1	135638	1	AF484556 Streptomy
38	338	14.9	300800	1	SC0939112 Streptomy
39	334	14.8	1329	6	AX433519 Sequence
40	329	14.6	302300	1	AF005034 Streptomy
41	328	14.5	2731	1	AF015225 Thermooc
42	328	14.5	6854	1	BSU39230 Bacillus sp
43	328	14.5	291000	1	SC0939105 Streptomy
44	327	14.5	300425	1	AP005044 Streptomy
45	324	14.3	301150	1	AP004602 Oceanobac

ALIGNMENTS

RESULT 1


```
Db 1261 CCATCTGGCCACACAGCGTTTCTCTCACTAGCTATCGTACAT 1299
RESULT 2
AB046402
LOCUS AB046402 1299 bp DNA linear BCT 16-JUN-2001
DEFINITION Bacillus sp. D6 PROA gene for protease, partial cds.
ACCESSION AB046402
VERSION AB046402.1 GI:12381936
KEYWORDS
SOURCE Bacillus sp. D6
ORGANISM Bacillus sp. D6
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (sites)
AUTHORS Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and
Horiuchi,K.
TITLE Novel oxidatively stable subtilisin-like serine proteases from
alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
evolutionary relationships
Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
JOURNAL
MEDLINE 20568675
PUBMED 11118284
REFERENCE 2 (bases 1 to 1299)
AUTHORS Saeki,K.
DIRECT SUBMISSION
Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
Biological Science Laboratory, Ichikaimachi Akabane 2605, Haga,
Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,
Tel:81-285-68-7400, Fax:81-285-68-7403)
FEATURES
Location/Qualifiers
1..1299
/organism="Bacillus sp. D6"
/mol_type="genomic DNA"
/strain="D6"
/db_xref="taxon:127889"
1..1299
/gene="PROA"
<1..>1299
/gene="PROA"
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAB21265.1"
/db_xref="GI:12381937"
/translation="NDVARGIVKDVQNNQGLYGQGVAVADTGLDTRNDSSMHE
AFRGKITALYALGRNNANDPNHGTHVAGSVLGNLKNKMAPQANLVFQSIIMDSGG
LGGLPSNLNTLPSQAWNAGARIHTNSWGA PVNGAVTANSROVDVVRNNDMTLPAAG
NEGNSGISAPGTAKNALTVGATENVPSSGIADNPNHIAQSSRGATDRIKPD
VTAPGTFILSARSSLAPSSFWANNYSKYAYGTSMTPIVAGNVAQLRHFYKNG
ITPKPSLIKAALIAGATDVGLGYPSSGDQGWKRVLDKSLNVAIVNEATALTIGQATY
SFQTAGKPLKISLVMTDAPGSTTASYTLVNDLDLIVITAPNGQKYGVNDFSYPYDNNW
DGRNNVENVFINAPQSGVTIEVQAYNPSPQRFSLAIVH"
ORIGIN
Alignment Scores:
Pred. No.: 2,34e-133 Length: 1299
Score: 2251.00 Matches: 430
Percent Similarity: 99.54% Conservative: 1
Best Local Similarity: 99.31% Mismatches: 2
Query Match: 99.56% Indels: 0
DB: 1 Gaps: 0
US-09-985-689A-5 (1-433) x AB046402 (1-1299)
QY 1 AsnAspValalaArgGlyIleVallysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 1 AATGATGTAGCAGAGGAATAGTAAAGACGAGCGTTGCACAAAACAAATTACGGACTATAT 60
QY 21 GlyGlnGlyClnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGCAAGGTCAAGTAGTTCAGTAGCGGATACCGGTTTATAGATACAGGTCTGTAACGATCT 120
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
121 TCTATGCATGAAGCAATCCGTGGGAAAATTACAGCTCTTTAGCCGTTAGGAGAACTAAC 180
61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
181 AATGCAAAATGATCCGAATGGGCATGGTACGCATGTAGCTGGTCTGTGCTTGGTAATGCT 240
81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
241 TTAATAAAGGAATGGCTCCGCAAGTAACTTAGTCTTCCCAATCTATTATGATAGCAGC 300
101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
301 GGGGATTAGGAGGATTACCATCGAATTTAAATACGTTATTATTAGTCAAGCTTGGAAATGCT 360
121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValaTyrThrAlaAsn 140
361 GCGCTAGTAATTCATACTAACTCTGGGGGGCCCAAGTAAATGGAGCGTACACTGCTAAC 420
141 SerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGly 160
421 TCGAGACAAGTGGATGAGTATGTTGMAACAATGATATGACGTACTTTTTCAGCTGGA 480
161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
481 AATGAAGGACCTAACTCTCGAAACAATTAGCGCTCCAGGACAGCGAAAATGCCATTACG 540
181 ValGlyValaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
541 GTCGGCGCAACGGAAACCTACCGACCAAGTTTGGTTCATTTGCAGATTAACCCCTAATCAT 600
201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
601 ATCGCAAAATTTTCATCGAGAGGAGTACGAGAGATGGAGCAATTAACACGACGCTAAC 660
221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
661 GCTCCTGGAAACATTTATATATATCAGCACGCTCTTCTTTAGCACCGACACTCTTCGTTGG 720
241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
721 GCGAATTTAAACAGTAGTATGCGTATATGGGCGGTACCTCTATGGCGACACCTATAGTT 780
261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
781 GCGGGGAATCTCGCGCAATTTACGCGAGCATTTTATAAAAAATAGAGAAATTCACCTAAA 840
281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyValaThrAspValGlyLeuGlyTyrPro 300
841 CCTTCCTTAATAAAGCTGCACATTATCGCTGGGGCTACTGATGTTGTTTAGGATATCCA 900
301 SerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
901 AGTGGTGACCAAGCTGGGGCGTGTACTTTAGATAAAATCGTTAAATAGGCGTATGTC 960
321 AsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAla 340
961 AATGAAGCAACTGCATTTAAACAACAGACAAAAGCAACGATTTCTGTTCCAAACGCAAGCG 1020
341 GlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer 360
1021 GGTAAACCATTAATAAATCTCGTTAGTATGGACAGATGCACCTGGAAGTACAAACAGCATCT 1080
361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
1081 TATACACTAGTTAATGATTTAGATCTAGTTATTACTGCTCCGATGGACAAAAATATGTA 1140
381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400
1141 GGTAAATCATTTAGTTATCTTATGATAATAATTTGGATGGTGGTCCGCAACAATTTGGAAC 1200
401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
1201 GTATTTAAACGCTCCGCAATCTGGAAACGCTATACAAATTTAGGTTCAAGCGGTATACGTT 1260
```

```

Qy 421 ProSerGlyProClnArgPheSerLeuAlaIleValHis 433
Db 1261 CCATCTGGACCACAGCGTTCTCCTAGTATCGTACAT 1299

RESULT 3
E03808
LOCUS      2218 bp      DNA      linear      PAT 29-SEP-1997
DEFINITION      DNA encoding alkaline protease Ya.
ACCESSION      E03808
VERSION      E03808.1 GI:2172022
KEYWORDS      JP 1992197182-A/1.
SOURCE      Bacillus sp.
ORGANISM      Bacillus sp.
REFERENCE      1 (bases 1 to 2218)
AUTHORS      Tobe, S., Odera, M. and Asai, Y.
TITLE      DNA CODING ALKALINE PROTEASE YA ENZYME AND PRODUCTION OF ALKALINE
JOURNAL      PROTEASE YA USING THE DNA
PATENT      JP 1992197182-A 1 16-JUL-1992;
LION CORP
COMMENT      OS      Bacillus sp.
PN      JP 1992197182-A/1
PD      16-JUL-1992
PE      28-NOV-1993 JP 1990327110
PI      TOSE SEIICHI, ODERA MOTOYASU, ASAI YOSHIO
PC      C12N15/57, C11D3/386, C12N9/54, (C12N15/57, C12R1:07), (C12N9/54,
PC      C12R1:07);
CC      strandedness: Double;
CC      topology: Linear;
CC      hypothetical: No;
CC      anti-sense: No;
CC      *source: strain=Y strain;
CC      *source: clone=pUB8A;
FH      Key
PH      Location/Qualifiers
FT      CDS      218..2125
FT      /product='precursor of alkaline protease Ya'
FT      sig_peptide      218..823
FT      mat_peptide      824..2122
FT      /product='Precursor of alkaline protease Ya'.

FEATURES
source
Location/Qualifiers
1..2218
/organism='Bacillus sp.'
/mol_type='genomic DNA'
/db_xref='taxon:1409'

ORIGIN

Alignment Scores:
Pred. No.:      1.06e-132      Length:      2218
Score:      2245.00      Matches:      429
Percent Similarity:      99.77%      Conservatives:      3
Best Local Similarity:      99.08%      Mismatches:      1
Query Match:      99.29%      Indels:      0
DB:      6      Gaps:      0

US-09-985-689A-5 (1-433) x E03808 (1-2218)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnTyrGlyLeuTyr 20
Db 824 AATGATGTACAGAGGATAGTAAAGCTGATGTTGCACAAAACAAATACCGATATAT 883

Qy 21 GlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 884 GGACAAGGTCAACTAGTTGTCAGTAGCGGACACAGGCTTAGATACAGGTCTACAGTAGT 943

Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 944 TCTATGATGAGGATTCGCGGGGAATACAGCTCTTTACCGTTAGGAAGAACTAAT 1003

Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 1004 AATCGAGTATCGAATGGGCATGGCAGCACATGATAGCAGGTTCTGCTGTTGTAATGCT 1063

Qy 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
Db 1064 TTTAATAAAGGAATGGCTCGCAGGTAAGTCTTCCATCTATTATGGATAGCAGC 1123

Qy 101 GlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
Db 1124 GGAAGGATTAGTGGCTTACCATCGAACTTAAATACGTTATTTAGTCAAGCTCGAATGCT 1183

Qy 121 GlyAlaArgIleHisThrAsnSerTTPGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
Db 1184 GGAAGCAGAATTCATCTAACTCTTGGGAGGCCCCAGTAAATGGAGCGTACACTGCTAAC 1243

Qy 141 SerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGly 160
Db 1244 TCGAGACACAGTGGATGATGATGTTGAAATAATGATATACGCTACTTTTTGCAGTGTG 1303

Qy 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 1304 AATGAAGTCTTAATTCAGGAACAATTAGTCTCCAGGTACAGCGAATAATGCTATTACG 1363

Qy 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
Db 1364 GTGGCGCAACGGAACAACTATCGCCAAAGCTTCGGTTCGATAGCAGATAACCCAAATCAT 1423

Qy 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
Db 1424 ATTGCACAATTTTCATCGAGAGGAGCTACGAGGATGGAGCAATTAAGCTGACGTAAACA 1483

Qy 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTTP 240
Db 1484 GCTCTCGAACATTTATTTATTCAGCACGCTTCTTCCTTAGCTCCAGACTCTTCGTTTGG 1543

Qy 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
Db 1544 GCGAATTAACAGATAAATACGCGTATATGGCGGACCTCCATCGGACACCATTTGTT 1603

Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
Db 1604 GCAGGGAATGCGCGCAATTCGTGAGCAATTTTATAAAAAATAGAGGTATTACTCCTAAG 1663

Qy 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
Db 1664 CCTTCCTTTAATAAAGCTGCACCTTATCGCTGGTGCTACTGATGTTGGTTTAGGATATCCT 1723

Qy 301 SerGlyAspGlnGlyTTPGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
Db 1724 AGTGTGACCAAGCTCGGGGGGTGTTACTCTAGATAAATCGTTAAATGTAGCGTATGTC 1783

Qy 321 AsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAla 340
Db 1784 AATGAAGCAACTGCATTAGCCACAGGACAAAAGCAACGATTTCGTTCCAAAGCACAGCG 1843

Qy 341 GlyLysProLeuLysIleSerLeuValTTPThrAspAlaProGlySerThrThrAlaSer 360
Db 1844 GGTAAACCTTTAAAAATCTCGTTAGTAGTACAGAGTCTCTCTGGAAGTACACTGCATCT 1903

Qy 361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
Db 1904 TATACACTAGTTAATGATTTAGATCTAGTTATTACTGCTCCGAATGGACAAAATATGTA 1963

Qy 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400
Db 1964 GGAATGATTTTAGTTATCTTATGATATAAATCTGGGATGGTGGCACCAATGTTTGAGAAC 2023

Qy 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
Db 2024 GTATTTATAACCGCTCCGCAATCTGGAACGTATATAATTCAGGTTCAAGCGTAAATGTA 2083

Qy 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 2084 CCATCTGCCACAGCGTTTCTCACTAGCTATCGTACAT 2122

```

```

RESULT 4
AB046404
LOCUS       AB046404.1 1299 bp DNA linear BCT 23-JAN-2001
DEFINITION Bacillus sp. Y PROC gene for protease, partial cds.
ACCESSION  AB046404
VERSION     AB046404.1
KEYWORDS    GT:12381940
SOURCE      Bacillus sp. Y
ORGANISM    Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE   1 (sites)
AUTHORS     Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and
            Horikoshi,K.
TITLE       Novel oxidatively stable subtilisin-like serine proteases from
            alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
            evolutionary relationships
JOURNAL     Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
MEDLINE     20568675
PUBMED      11118284
REFERENCE   2 (bases 1 to 1299)
AUTHORS     Saeki,K.
TITLE       Direct Submission
SUBMITTED   (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
            Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
            Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,
            Tel:81-285-68-7400, Fax:81-285-68-7403)
FEATURES             Location/Qualifiers
     source            1..1299
                     /organism="Bacillus sp. Y"
                     /mol_type="genomic DNA"
                     /strain="Y"
     gene              1..1299
                     /db_xref="taxon:133779"
     CDS               1..1299
                     /gene="PROC"
                     <1..>1299
                     /gene="PROC"
                     /evidence=experimental
                     /transl_table=1
                     /product="protease"
                     /protein_id="BAB21267.1"
                     /db_xref="GI:12381941"
                     /translation="NDVARGIVKADVAQNNTGLVGGQGVVAVADTGLDTRNDSSMHE
                     AFRGKITLALYALGRNTNADPNHGHTVAGSVLGNALNKGAPQANLIVFQSIMDSSGG
                     LGLEPSNLNTLFSQAWNAGARIHTNSWGPVNGAVTANSROVDYVRNNDMTLVFAAG
                     NEGPSNTISAPGAKNAITVGATENYRSPFGSLADNPNIHQFSSRGATDRGRIKPD
                     VTPGTFILSARSLAPDSSFWANYSKYVAMGTSMTPIVAGNVACLREHFKNKG
                     ITPKSLIKALIAAGATDVLGYPNGDQMGWVTLNKLNAVYNEALATLGOKATY
                     SFQAQAGPLKISLWTDAPGSTTASYIIVNDLDELIVITAPNGQKIVGNDFFSYPDNNW
                     DGRNVENVFINAPSGTIIIEVQAINVPSGPQRPFLAIVH"
ORIGIN
Alignment Scores:
Pred. No.:      1,16e-132      Length:      1299
Score:          2240.00      Matches:     428
Percent Similarity: 99.7%      Conservative: 4
Best Local Similarity: 98.8%      Mismatches: 1
Query Match:     99.0%      Indels:      0
DB:              1              Gaps:       0

US-09-985-689A-5 (1-433) x AB046404 (1-1299)

Qy      1  AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db      1  AATGATGATGACGAGGATAGTAAAGCTGATGTTGCACAAAACAATTACGGATTATAT 60
Qy      21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyValArgAsnAspSer 40
Db      61 GCGACAGGTCAAGTAGTGTGACATAGCGACACACAGCTTAGATACAGGTTCGTAAACGATAGT 120
Qy      41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db      121 TCATGTCATGAAGACATTCGCGGGGAAATCACAGCTCTTTACGCGTTAGGAAGAACTAAT 180
61  AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
181  AATGCGAGTGATCCGAATGGCATGGCACACATGTAGCAGGTCTCTACTTGTGTAATGCT 240
81  LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
241  TTAATAAAGGAATGGCTCGCAAGTAATCTTAGTCTTCCAATCTATTATGGATAGCAGC 300
101  GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAla 120
301  GCGAGGATAGTGGCTTACCATCGAATTAATACGTTATTTAGTCAAGCTTGGAAATGCT 360
121  GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
361  GCGACGAAGATTCTATCTAATCTTGGGAGGCCCAAGTAAATGGAGCGTACACTGTCTAAC 420
141  SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
421  TCGAGACAGTGGATGATATGTTCCAAATATGATATGACGTACTTTTTCGAGCTGGT 480
161  AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
481  AATGAAGGTCTTAATTCAGGAACAATTAGTCTCCAGGTACACGCAAAAATGCTATTACG 540
181  ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
541  GTCGGCGCAACGGAAAACATATCGCCAGCTTCGGTTCGATAGCAGATACACCAATCAT 600
201  IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
601  ATTGCACAATTTTCATCGAGGAGCTACGAGGATGGACGAATTAAGCTGACGTAAACA 660
221  AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
661  GCTCCTCGGAACATTTATTTATCAGCACGTTCTTCTTAGCTCCAGACTCTTCGTTTGG 720
241  AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
721  GCGAATTAATAACAGTAATAACGCGTATATGGCGGTACCTCCATGCGCACACCTATTGTT 780
261  AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
781  GCGAGGAATTCGCGCAATTTACGTGAGCATTTTATAAAAAATAGAGGTATTACTCTCAG 840
281  ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
841  CCTTCTTTAATAAAGCTGCACCTTATCGCTGGTGTCTACTGATGTTGGTTTAGGATATCT 900
301  SerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
901  AATGCTGACCAAGCTGGGGCGGTGTTACTCTAAATAAATCGTTAAATAGCGTATGTC 960
321  AsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAla 340
961  AATGAAGCAACTGCATTTAGCCACAGACAAAAGAACGCTATTTCGTTCCAGCACAGCG 1020
341  GlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer 360
1021  GGTAAACCTTTAAAAATCTCGTTAGTATGACACAGATGCTCCTGGAAGTACAACTGCATCT 1080
361  TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
1081  TATACACTAGTTAATGATTAGTATCTAGTATTACTGCTCCGAATGGACAAAATATGTA 1140
381  GlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyValArgAsnAsnValGluAsn 400
1141  GGAATGATTTTAGTTATCTTATGATATAATACTGGATGGTGGCAACAATGTTGAGAAC 1200
401  ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnIleTyrAsnVal 420
1201  GTATTATTAACGCTCCGAATCTGGAACTGATATAAATTTAGGTTTCAGCGTATATGTA 1260

```



```
Qy 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
|||||
Db 1261 CCATCTGGCCACAGCGTTTCTCACTAGCTATCGTACAT 1299

RESULT 5
AR069954
LOCUS AR069954 3003 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 41 from patent US 5891701.
ACCESSION AR069954
VERSION AR069954.1 GI:7220842
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1. (bases 1 to 3003)
AUTHORS Sloma, A. and Christianson, L.
TITLE Nucleic acids encoding a polypeptide having protease activity
JOURNAL Patent: US 5891701-A 41 06-APR-1999;
FEATURES
LOCATION/Qualifiers
1..3003
source /organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.: 7,24e-120 Length: 3003
Score: 2044.00 Matches: 386
Percent Similarity: 94.69% Conservative: 24
Best Local Similarity: 89.15% Mismatches: 23
Query Match: 90.40% Indels: 0
DB: 6 Gaps: 0

US-09-985-689A-5 (1-433) x AR069954 (1-3003)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaClnAsnAsnTyrGlyLeuTyr 20
|||||
Db 1470 AATGACGTGCCCGTGGCATTTGTGAAGCAGACGTGCACAAAATAACTTTGGCTTATAT 1529

Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
|||||
Db 1530 GGACAAGGACAGATTGTAGCAGTTGCTGATCTGGCTTGTATACAGGAAGATGACAT 1589

Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
|||||
Db 1590 TCGATGATGAAGCATTCGCGGTGAAGTTACCGCACTATATACCTGCGCAGACGAAT 1649

Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyValAsnAla 80
|||||
Db 1650 AACCCCAATGATCCAAATGGACATGGAAACCATGTTGCTGGATCTGTGTAGGAATGCT 1709

Qy 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
|||||
Db 1710 ACAATAAAGGATGGCAGCCGCAAGCAATCTAGTCTTCAATCTATTATGGATAGTGT 1769

Qy 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTTPAsnAla 120
|||||
Db 1770 GGAGGCTGGGAGACACCTCTGCTAACTCAACATATTTCAGTCAAGCATATAGTCT 1829

Qy 121 GlyAlaArgIleHisThrAsnSerTTPGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
|||||
Db 1830 GGAGCGAGAAATTCATACGAATTCATGGGGCTCCAGTAAACGGTGCCTATACGACAGAC 1889

Qy 141 SerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGly 160
|||||
Db 1890 TCTCGAATGTGTATGATATTGTGAGAAAAAATGATATGACGATCTTTTGGCGCGGA 1949

Qy 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
|||||
Db 1950 AATGAGGACACGAGTAGCGGTACAACTCAGTGCACCAAGCAAGCAAAATGCGATTACA 2009

Qy 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
|||||
Db 2010 GTTGGGCGACCGAAACCTACGTCCAAAGCTTCGGATCTTATCGGATAATATTAACCAT 2069
```

```
Qy 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
|||||
Db 2070 GTTGTCTCAATTCTCTTCAAGAGTCTCTACTAGAGATGAGCGTATTAAAGCCGAGCTCATG 2129

Qy 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTTP 240
|||||
Db 2130 GCACGAGTACGTATATCTCTCTGTAGATCATCATTTAGTCCAGATTCCTCATTTCTGG 2189

Qy 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
|||||
Db 2190 GCAAAACCATGATAGTAATATGCTACATGGTGGTACTTCTATGGCTACTCCAAATGTA 2249

Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
|||||
Db 2250 GCAGGTAATGTGCACAAATTAAGGAGCATTTTGTGAANAATAGAGGGGTAACTCCTAAG 2309

Qy 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
|||||
Db 2310 CTTCCCTTTTAAAGGCTGCTTTAATTGCAGGTGCTCGGATGTTGGACTTGGCTTTCCA 2369

Qy 301 SerGlyAspGlnGlyTTPGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
|||||
Db 2370 AATGTAACCAAGGATGGGAAGAGTAACGTTAGATAAATCCCTAAATGTGCGATTTGTG 2429

Qy 321 AsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAla 340
|||||
Db 2430 AATGAAGCAGCCTTTATCAACAGTCMAAAGCAACATATTCGTTTACGGCTCAAGCT 2489

Qy 341 GlyLysProLeuLysIleSerLeuValTTPThrAspAlaProGlySerThrThrAlaSer 360
|||||
Db 2490 GGTAACCCCTTAAATAATATCATCTTGTGTGTCAGATGCCAGGTAGCAGCGCATCA 2549

Qy 361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
|||||
Db 2550 CTACTTTAGTGAATGATTAGACTTATGATCTACTGCACCAATAGCAATTAATAGCTC 2609

Qy 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTTPAspGlyArgAsnAsnValGluAsn 400
|||||
Db 2610 GGAAATGACTTACAGCACCGCTATGATAACAATTTGGATGGCAGAAACAACGCGAAAAAT 2659

Qy 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
|||||
Db 2670 GTGTTTATCAATGCTCTCTCAAGCGGAACGATATACGTGGAAGTGCAGGCTTCAATGTA 2729

Qy 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
|||||
Db 2730 CCAGTAAGTCGCAAAACCTTTTCTTTAGCGATTGTACAT 2768

RESULT 6
BD062155 3003 bp DNA linear PAT 27-AUG-2002
LOCUS BD062155
DEFINITION Nucleic acids encoding a polypeptide having protease activity.
ACCESSION BD062155
VERSION BD062155.1 GI:23607760
KEYWORDS JP 2001514529-A/39.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
1. (bases 1 to 3003)
AUTHORS Sloma, A. and Christianson, L.
TITLE Nucleic acids encoding a polypeptide having protease activity
JOURNAL Patent: JP 2001514529-A 39 11-SEP-2001;
COMMENT
NOVO NORDISK BIOTECH INC
PN JP 2001514529-A/39
PD 11-SEP-2001
PF 09-JUN-1998 JP 1999503145
PI 12-JUN-1997 US 08/873479
PT ALAN SLOMA, LYNNE CHRISTIANSON
PC C12N15/57, C12N15/75, C12N9/54, C12K14/00
CC Strandedness: Single;
Topology: Linear;
```

FEATURES

FH	Key	Location/Qualifiers	Location/Qualifiers
FEATURES	source	1..3003	
		/organism="Homo sapiens"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:9606"	
ORIGIN			
Alignment Scores:			
Pred. No.:	7,24e-120	Length:	3003
Score:	2044.00	Matches:	386
Percent Similarity:	94.69%	Conservative:	24
Best Local Similarity:	89.15%	Mismatches:	23
Query Match:	90.40%	Indels:	0
DB:	6	Gaps:	0
US-09-985-689A-5 (1-433) x BD062155 (1-3003)			
QY	1	AsnAspValAlaAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr	20
Db	1470	AATGACGTGGCCGTGGCATTGTGAAGCAGACAGCTCGCACAAATTAACCTTTGGCTTATAT	1529
QY	21	GlyGlnGlyValValAlaAlaAspThrGlyLeuAspThrGlyValArgAsnAspSer	40
Db	1530	GGCAAGGACAGATTGTAGCAGTTGCTGATCTGGCTTGATACAGAGAAATGACAGT	1589
QY	41	SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn	60
Db	1590	TCGATGCATGAAGCATTCGCGGTAAAGATTACCGCACTATATGCTGGCGAGAACGAT	1649
QY	61	AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla	80
Db	1650	AACGCCAATGATCCAAATGGACATGGAAACCCATGTTGCTGGATCTGTGTAGGAATGCT	1709
QY	81	LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer	100
Db	1710	ACAAATAAGGATGGCAGCGCAAGCAATCTAGTCTTTCAATCTATTATGATAGTGT	1769
QY	101	GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAla	120
Db	1770	GGAGGGCTGGAGGACTACTGCTAATCTCAACACATTATTCAGTCAAGCATATAGTGT	1829
QY	121	GlyAlaArgIleHisThrAsnSerTyrGlyValAlaProValAsnGlyAlaTyrThrAlaAsn	140
Db	1830	GGAGCGAGATTTCATACGAATTCATGGGGGCTCCAGTAAACGGTCCCTATACGACAGAC	1889
QY	141	SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly	160
Db	1890	TCTCGAAATGTTGATGATTATGTGAGAAATAATGATATGACGATCTTTTTCGGCCGGA	1949
QY	161	AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr	180
Db	1950	AATGAGGACCGACGTAGCGGTACAAATCAGTGCACCGAGAACACAAATAATGCGATTACA	2009
QY	181	ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis	200
Db	2010	GTTTGGGCAACCGAAAACTACGTCCAGCTTCGATCTTATCGGATATATATTAACCAT	2069
QY	201	IleAlaGlnPheSerSerArgGlyValAlaThrArgAspGlyArgIleLysProAspValThr	220
Db	2070	GTTGCTCAATCTCTTCAGAGGTCCTCTAGATGAGATGAGTATTAGCCGCGACGTCATG	2129
QY	221	AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr	240
Db	2130	GCACCGGTACGTATATCTCTCTGTAGATCATCATCTAGCTCCAGATTCCTCATCTGG	2189
QY	241	AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal	260
Db	2190	GCAACCATGATAGTAATATGCGCTCATGCGGTGGTACTCTTATGGCTACTCCCAATTGTA	2249
QY	261	AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys	280
Db	2250	GCAGGTAATGTTGCACAAATTAAGGGAGCATTTTGTGAAAAAATAGAGGGTAACTCCTAAG	2309

QY	281	ProSerLeuIleLysAlaAlaIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro	300
Db	2310	CCTTCCCTTTTAAAGAGCTGCTTTAATGTCAGGTCTCGGATGTTGGACTTGGCTTTCCCA	2369
QY	301	SerGlyAspGlnGlyTyrGlyValThrLeuAspLysSerLeuAsnValAlaTyrVal	320
Db	2370	AATGTTAACCAAGGATGGGAAGAGTAGTAAATCCCTAAATGTCGCATTTGTG	2429
QY	321	AsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAla	340
Db	2430	AATGAAACGAGCCCTTTATCAACAGTCAAAAGCAACATATTCGTTTACGGCTCAAGCT	2489
QY	341	GlyLysProLeuLysIleSerLeuValThrThrAspAlaProGlySerThrThrAlaSer	360
Db	2490	GGTAAACCCCTTAAAAATATCACTGTTGTTGGTCAGATGCACAGGTAGCAGCGCATCA	2549
QY	361	TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal	380
Db	2550	CTAACTTTAGTGAATGATTTAGACTTAGTAATCACTGCACCAATAGGAATAATAGCTC	2609
QY	381	GlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyValGlnAsnValGluAsn	400
Db	2610	GGAAATGACTTTTACAGCACCGTATGATACAAATTTGGGATGCGAGAACACGTCGAAAT	2669
QY	401	ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal	420
Db	2670	GTGTTTATCAATGCTCCTCAAGCGGAACGTATACAGTCGAAGTGCAGGCTTACAATGTA	2729
QY	421	ProSerGlyProGlnArgPheSerLeuAlaIleValHis	433
Db	2730	CCAGTAAGTCCGCAAAACCTTTTCTTTAGCGATTGTACAT	2768
RESULT 7			
LOCUS	AB046406	1302 bp	DNA linear BCT 23-JAN-2001
DEFINITION	Bacillus sp. NV1	PROE gene for protease, partial cds.	
ACCESSION	AB046406		
VERSION	AB046406.1	GI:12381944	
KEYWORDS			
SOURCE	Bacillus sp. NV1		
ORGANISM	Bacillus sp. NV1		
REFERENCE	1 (sites)	Bacteri; Firmicutes; Bacillales; Bacillaceae; Bacillus.	
AUTHORS	Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and Horikoshi, K.		
TITLE	Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic Bacillus spp.: enzymatic properties, sequences, and evolutionary relationships		
JOURNAL	Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)		
MEDLINE	20568675		
PUBMED	11118284		
REFERENCE	2 (bases 1 to 1302)		
AUTHORS	Saeki, K.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-JUL-2000) Katehisa Saeki, Kao corporation, Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga, Tochigi 321-3497, Japan (E-mail:387185@kstanet.kao.co.jp, Tel:81-285-68-7400, Fax:81-285-68-7403)		
FEATURES	Location/Qualifiers		
source	1..1302	/organism="Bacillus sp. NV1"	
		/mol_type="genomic DNA"	
		/strain="NV1"	
		/db_xref="taxon:133781"	
	1..1302	/genes="PROE"	
gene	<1..>1302	/gene="PROE"	
CDS		/codon_start=1	
		/transl_table=11	
		/product="protease"	
		/protein_id="BAB21269.1"	

RESULT 7

AB046406	1302 bp	DNA	linear	BCT 23-JAN-2001
LOCUS	AB046406			
DEFINITION	Bacillus sp. NV1 PROBE gene for protease, partial cds.			
ACCESSION	AB046406			
VERSION	AB046406.1 GI:12381944			
KEYWORDS				
SOURCE	Bacillus sp. NV1			
ORGANISM	Bacillus sp. NV1			
REFERENCE	1 (sites)			
AUTHORS	Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and Horikoshi,K.			
TITLE	Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic Bacillus spp.: enzymatic properties, sequences, and evolutionary relationships			
JOURNAL	Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)			
MEDLINE	20568675			
PUBMED	11118284			
REFERENCE	2 (bases 1 to 1302)			
AUTHORS	Saeki,K.			
TITLE	Direct Submission			
JOURNAL	Submitted (20-JUN-2000) Katsuhisa Saeki, Kao corporation, Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga, Tochigi 321-3497, Japan (E-mail:387185@kastonet.kao.co.jp, Tel:81-285-68-7400, Fax:81-285-68-7403)			
FEATURES	Location/Qualifiers			
Source	1..1302			
	/organism="Bacillus sp. NV1"			
	/mol_type="genomic DNA"			
	/strain="NV1"			
	/db_xref="taxon:133781"			
gene	1..1302			
	/gene="PROE"			
CDS	<1..>1302			
	/gene="PROE"			
	/codon_start=1			
	/transl_table=11			
	/product="protease"			
	/protein_id="BAB21269.1"			

/db_xref="GI:12381945"
/translation="NDVARGIVKADVAQSSVGLYGGGVAVADTGLDGRDSSMHE
AFRGKITALIYLRNNANDPNKHGTHVAGSVLNGTNGKMAPOANLVPOVMSNG
GLGLPSNVSTLFSQYAKNITVNSGAPVNGAITTDSRVDDIYVRKNDVLFPA
GNEGNGGTISAPGTAKNAITVATENRPSFGYADNINHVQPSFGPKTDGRKIP
DYMAPGTFLISARSSLPADSPFWANHSKYMGSTWATPIVAGNVAQLREHFKNR
GITPKPSLKAALIAAGTIDIGLSPSGNGVRVLDKSLNVAFNYSSTLSTNQAT
YSEFTAGSGKPLKISLVSDAPASTSASVTLNVDLIVITAPNGTKYVGNDFTPYDNN
WDGRNNVENVFNAPOSGTYTVEVQAYNVPGQPAFLAIVN"

ORIGIN

Alignment Scores: 4.57e-119 Length: 1302
Pred. No.: 2024.50 Matches: 384
Score: 94.93% Conservatives: 28
Percent Similarity: 94.93% Mismatches: 21
Best Local Similarity: 88.48% Indels: 1
Query Match: 89.54% Gaps: 1
DB: 1

US-09-985-689A-5 (1-433) x AB046406 (1-1302)

QY 1 AsnAspValAlaAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 1 AAGCATGTTGCCAGAGCATGTAAAGCCGATGTTGCCAGAGCATGATGTTATAT 60
QY 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGGCAAGGGCAAGTGTTCAGTAGCGGATACCGGACTGGATACAGGGCGTAATGACAGC 120
QY 41 SerMetHisGluAlaPheAeGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCGATGATGAGGCTTCCGAGGAGAAAGATTACCGGATATATGCCCTTGGAGACAAAC 180
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 181 AACGCCAATGATCCAAACGGACACGGGACGCATGTTCCCGGATCTGTTTAGGAAACGGT 240
QY 81 Leu---AsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 241 ACTTCGATAAAGGATGGCTTCCCAAGCTNACTTAGTTTCCNATCTGTATGAGCAGC 300
QY 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 301 AATGGCGACTTGGCGGACTGCCITCCAAATGTAAGTACATATTATTCAGCCAGGCATATAGT 360
QY 120 AlaGlyAlaAeGlyIleHisThrAsnSerTrpGlyAlaProValAlaGlnGlyValAlaTyrThrAla 139
Db 361 GCCGTGCCAGAAATCCATACGAATCTATGGGAGCGCTGTAAACGGAGCGTACACTACT 420
QY 140 AsnSerArgGlnValAlaSpGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 421 GATTCAGAAACGTAGACGATTATGTTCTAAATAATGATATGGCGGTCTCTTTTGCAGCG 480
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 481 GGTAAACGAGAGGCCAAATGGCGGCAATCATAGTCTCTGTATCCCGGAGAAATGCTATC 540
QY 180 ThrValGlyValAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
Db 541 ACAGTAGGGGCACAGAAACCTCGCCCAAGCTTTGCATCTTATGTCGACAAACATCAAT 600
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 601 CATGTAGCACAGTCTTCTCCCGGACCTCAAAAGGATGACGATCAAAACCGGACGTA 660
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
Db 661 ATGGCACACGAGAACATTATTTATTCGCAAGATCTCTTTGGCTCCGAGCATCTTCATTC 720
QY 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
Db 721 TGGGCAACCATGACAGCAAAATATGTTATATGCGTGGAGAACATCATCCATCGGACCGGATT 780

QY 260 ValAlaGlyAsnValAlaGlnLeuAlaArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 781 GTAGCTGGTAACCTTGACAGTGTACGTGAACATTTTCATCAAAACAGAGGAATCACTCTC 840
QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyValaThrAspValGlyLeuGlyTyr 299
Db 841 AACCATCTCTTGTGAAGAGAGCTCTTATTGCGGAGCACTGATATCGTCTTGGCTAT 900
QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 901 CCGAGTGGAAACCAAGCATGGGAAGAGTAGTAACATTGACAAAGTCATTAATGTAGCTTTC 960
QY 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
Db 961 GTAAATGAACACAGCTCTTTATCTACTACCAAAAGGCTACGTATTCATTACTGCACAA 1020
QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAla 359
Db 1021 TCAGGCAAAACCTTGAAGATTTTCATTGTTGGTCTGATGACCGGCAAGTACTTCGCA 1080
QY 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1081 TCGGTTCATTTGGTGAATGATCTGGATCTGGTATTACAGCTCCAAATGGAACAAAGTAT 1140
QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
Db 1141 GTTGGAAACGACTTTACTGCTCCCTATGATAATACTAGGATGACGCTAACATGTAGAG 1200
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1201 AACGTGTTTATCAATGCTCCGAAAGCGAAGCATATACAGTTGAGGTACAGGCTTACAT 1260
QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1261 GTTCCACAGGGCCGCGCGGCTTTCTTTGGCTATTGTGAAC 1302

RESULT 8

AB084155
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

FEATURES
source

CDS

1..1923
/organism="Bacillus sp. KSM-9865"
/mol_type="genomic DNA"
/strain="KSM-9865"
/db_xref="taxon:192495"
1..1923
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAC82522.1"
/db_xref="GI:34392387"
/translation="NRKKKVFSLVLSAAIILSTVALSNPSAGARNFLDPKGIQTI
TPAKGSKOGTGAALFLVSENVKLPKGLQKLETVPAKKLPIVQFNPGPILEETKQ
OLEKTKAKILDIPYAVIYVEGDKVSTIEDVESVEPLPIYRIDPOLFTKGAS
ELVKAVALDITNKNKEVQLRGIEQIAQFATSDNVLITAKPEYKWNVDYARGIVKADV

AB084155 1923 bp DNA linear BCT 02-SEP-2003
Bacillus sp. KSM-9865 gene for protease, complete cds.

AB084155.1 GI:34392386

Bacillus sp. KSM-9865
Bacillus sp. KSM-9865

Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

Okuda, M., Saeki, K. and Kobayashi, T.

Bacillus sp. KSM-9865 protease gene

Published Only in Database (2003)

2 (bases 1 to 1923)

Okuda, M., Saeki, K. and Kobayashi, T.

Direct Submission

Submitted (18-APR-2002) Mitsuyoshi Okuda, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: okuda.mitsuyoshi@kao.co.jp,

Tel:81-285-68-7543, Fax:81-285-68-7547)

Location/Qualifiers

AOSSYGLYGOGIIVAVADTCLDTGRNDSNMHEAFRCIKITALLYALQTSQVANNANDTNGHGT
HVAGSVLNGSTKNGKAPQANLVFQSIMDSGGGLGGLPSNLQTLRSQVAGARITHYN
SWGAANVAGTNNVDDIVRNKNDMTILFAENGFGNGTISAPGTAKNALIVGATE
NLRFPSGADNINHVAFSSRGTDKRIKPDVMAFGPTFILSARSSLPDPSFWANH
DSKYAYMGSTSMATPIVAGNVAQREHFVKNRGITPKPSLLKALITAGAADIGLGYPN
GNQGWGRVTLDKSLINWAYNESSLSSTSQKATYSTFTATGPKLISLVMSDAPASTTA
SVTLVNDLDLVIITAPNGTQVGVNDFTSPYNNWMDGRNVENFINAPQSGTGTIEVQA
YNYVPVGPQTSALAIVN"

ORIGIN

Alignment Scores:

Pred. No.: 1.53e-117 Length: 1923
Score: 2003.50 Matches: 382
Percent Similarity: 94.24% Conservative: 27
Best Local Similarity: 88.02% Mismatches: 24
Query Match: 88.61% Indels: 1
DB: 1 Gaps: 1

US-09-985-689A-5 (1-433) x AB084155 (1-1923)

QY 1 AsnAspValAlaAAGGlyIleValLysAlaAspValAlaAlaGlnAsnAsnTyGlyLeuTy 20
DB 619 AATGATGTTGACGTGGAATGTCAAAGCGGATGTGGCGCAGACGACGTACGGGTGTAT 678
QY 21 GlyGlnGlyGlnValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 679 GGCACAGACAGATCGTAGCGGTTGCCGATACAGGGCTTGATACAGGTCCCAATGACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyAlaLeuGlyArgThrAsn 60
DB 739 TCGATGATGATGAGCTTCCGGGGGAAATTAATCTGATATATGATTCGATTCGGACGACGAT 798
QY 61 AsnAlaAsnAppProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
DB 799 AATGCCAATGATACGAATGGTCATGGTACGCATGTGGCTCGGTATTAGGAAACGGC 858
QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
DB 859 TCCACTAATAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCAACTATCATGATGATGC 918
QY 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
DB 919 GGTGGGGACATTGGAGGACTACCTCGAATCTGCAAACTTATTTCAGCAAGCATACAGT 978
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrAla 139
DB 979 GCTGTGCCGAATTCATACAACTCTCGGGAGCAGCAGTGAATGGGCTTACACAA 1038
QY 140 AsnSerArgGlnValAspGlyLysValArgAsnAsnAspMetThrValLeuPheAlaAla 159
DB 1039 GATTCAGAAATGTGATGACTAATGTGGCAAAATGATATGACATCTTTTCGTGCC 1098
QY 160 GlyAsnGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
DB 1099 GCGAATGAAGGACCGAAGCGGGAACCATCATGTCACCAAGCAGCAGCTTAAATA 1158
QY 180 ThrValGlyAlaThrGluAsnTyArgProSerPheGlySerLeuAlaAspAsnProAsn 199
DB 1159 ACATGCGGAGCTACGGAACCTCGCCAGCTTCGGGTCTTATGCGGCAATATCAAC 1218
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
DB 1219 CATGTGGCAGCTTCTCTTCGTGGACCGCAAAAGGATGGACGATCAAAACCGGATGC 1278
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
DB 1279 ATGGCACCGGGAACGTTCTATATATCATATCAGCAAGATCTTCTGTGACCGGATCTCTCTC 1338
QY 240 TrpAlaAsnTyArgSerLysTyAlaTyMetGlyThrSerMetAlaThrProIle 259
DB 1339 TGGGCGAACCATGACAGTAAATATGCATACATGGGTGGAACTCCATGCTGCTACACCGATC 1398
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279

DB 1399 GTTCTGGAACGTGGCACACGCTTCGTGAGCATTTTGTGAAAAACAGAGGCATCACACCA 1458
QY 280 LysProSerLeuIleLysAlaAlaIleuIleAlaGlyAlaThrAspValGlyLeuGlyTy 299
DB 1459 AAGCTTCTTATTAAAGCGGCACTGATTCGCCGTGCGAGACATCGGCTTGGCTAC 1518
QY 300 ProSerGlyAspGlnGlyTyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTy 319
DB 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATTTGGATAAATCCCTAAACGTTGCCTAT 1578
QY 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTySerPheGlnAlaGln 339
DB 1579 GTGAACGAGTCCAGTTCCTATCCACCACCAAGCGAGCTACTCGTTTACTGCTACT 1638
QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrAla 359
DB 1639 GCCGCAAGCCTTTGAAGATCTCCTCGTATGCTGATGCCCTGGCAGCACAACTGCT 1698
QY 360 SerTyThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTy 379
DB 1699 TCCGTAAACGCTTGTCAATGACCTGGACCTTGTCAATACCGCTCCAAATGGCACCAAT 1758
QY 380 ValGlyAsnAspPheSerTy-ProTyArgAsnAsnTrpAspGlyArgAsnValGlu 399
DB 1759 GTTGAAGATGACTTACTTCGCCATACATAATACTGGATGGCCGCAATAACGTAGAA 1818
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyThrIleGluValGlnAlaTyAsn 419
DB 1819 AATGTATTATTATATCGCCCAAAAGCGGAGCTATACAAATTGAGGTACAGCTTATAAC 1878
QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
DB 1879 GTACCGTTGGACCAACACCTTCTCGTTGGCAATTGTGAAC 1920
RESULT 9
LOCUS AB046403 1920 bp DNA linear BCT 10-MAY-2002
DEFINITION Bacillus sp. 9860 PROA gene for protease, complete cds.
ACCESSION AB046403
VERSION AB046403.2 GI:20521152
KEYWORDS Bacillus sp. 9860
SOURCE Bacillus sp. 9860
ORGANISM Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1
AUTHORS Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and Horikoshi, K.
TITLE Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic Bacillus spp.: enzymatic properties, sequences, and evolutionary relationships
JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
MEDLINE 20568675
PubMed 11118284
REFERENCE 2 (bases 1 to 1920)
AUTHORS Saeki, K.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation, Biological Science Laboratory, Ichikaimachi Akabane 2606, Hagi, Tochigi 321-3497, Japan (E-mail: 387185@kamatanet.kao.co.jp, Tel: 81-285-68-7400, Fax: 81-285-68-7403)
COMMENT On May 9, 2002 this sequence version replaced gi:12381938.
FEATURES
source
1. .1920
/organism="Bacillus sp. 9860"
/mol_type="genomic DNA"
/strain="9860"
/db_xref="taxon:133778"
1. .1920
/gene="PROA"
1. .1920
/gene="PROA"
/codon_start=1
gene
CDS

```
/transl_table=11
/product="protease"
/protein_id="BA821266_2"
/db_xref="GI:20521153"
/translation="MRKKKVFLLSVLSAAIILSTVALNPPSAGDARTFDLPKGIQTFT
DVSFSGKORQGAAPLVESSENVKLLKGLKLETVFANNKHLIVQNGFLEIKKK
LETTGAKLIDYIPDYAYIVEGVDQSVRSIEHVESVEPYLPKIDPDQFTKGAST
LVKALALDTQKNKEVQVQRIEIEIAQYVASNDVHYITAKPEYKVMNDVARGIKADVA
QSSVGLNGQVQVAVADTGLDTRNDSSMHEAFRGKITALYALGRTNANDTNGHGH
VAGSVLGNKGMKNGMAPOANLVFQSIMDSGGLGLPSNLOTLFQAQPSAGARIHTNS
WGAANGVATYDTSRNVVDYVKNQDMTILPAAGNERPNGGTISAPGTAKNAITVGTATN
LRPSGSDADNINVAQSSRGRTQKRIKEDVWAPGYIILSASSSLAPDSSFWANDH
SKYAYNGGTSMAPIVAGNVAQLRHFVKNRGITIPKPSLLKALAGADUUGLGYPNG
NQGWGRVLDLKSNAVYVNESSALTSQATVYFTATAGKPKIKLSLWMSDAPASITTS
VTLVNDLDIVITAPNGTRYVGNDFAPFEDNNWGDGRNVNFIINSPOSGTYTIEVQAY
NVVPGPQNPFLAIVN"
ORIGIN
Alignment Scores:
Pred. No.: 1,77e-117 Length: 1920
Score: 2002.50 Matches: 383
Percent Similarity: 94.24% Conservative: 26
Best Local Similarity: 88.25% Mismatches: 24
Query Match: 88.57% Indels: 1
DB: 1
Gaps: 1
US-09-985-689A-5 (1-433) x AB046403 (1-1920)
QY 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnAsnAsnTrpGlyLeuTyr 20
Db 616 AATGATGTGCCAGAGGTAATGTCAAGCGGATGTGGCACAGACGACTACGGTTTGTAT 675
QY 21 GlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGCACAGCGCAGATGTGCGAGTGTGCGATCTGGATTGGATACAGAGAAACACGACGT 735
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCATGAGCCITCCGCGGTAAATAACAGCACTATATGCACTGGGTGGAGCAAT 795
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 796 AATCGCAATGATACGACCGTGTGTTACCATGTCAGTGGAGTTCGGTATTAGGAATGCG 855
QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 856 GCAACGATTAAGGATGGCACTCAAGCGAATCTGGTTTTCATCCATCATGTGATACG 915
QY 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 916 AGTGTGGTGGCTTGGAGGCTTGCCTTCCAACTCGCAACACCTATTTCAGCCAAAGCAATTCAGT 975
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAla 139
Db 976 GCAGTGCACAGATTCATACAACTCTCGGGGCGACGGTGAATGGGCTACACGACA 1035
QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 1036 GATTCCAGAAATGTGGATGACTATGTAAAGGAAATATGATGACGATTCCTTTTCGGCGGT 1095
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1096 GGGAAATGAAGCCGACCGCGGTACCATGATGACCTGTCAGCTGGTACGGTAAACGCCATA 1155
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
Db 1156 ACAGTGCAGCAACCGAAACCTCGCTCAAGCTTCGGTTCCTATGTCAGATAATATTAAAC 1215
QY 200 HisIleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 1216 CACGTTGCACAGTCTCTCTCCCGTGGCCCGCAAGAGATGGCGAAATCAAGCCTGATGTC 1275
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
```

```
1276 ATGGCCCGGAGGACATACATATTTTATCAGCAAGATCTTCTTGGACCCGATTCCTCTTC 1335
240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
1336 TGGCGCAATCATGACAGCAAAATATGCTATATGGGTGGAAACGTCCTCATGCAACACCGATT 1395
260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
1396 GTTTCGGGGGAATGTTCCACAGCTCCGTGAGCATTTTGTGAAAAATAGAGAACTACTCTCT 1455
280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
1456 AAGCTTCCCTATTGAAAGCAGCTTTGATTGAGGTGCTGCTGATGTTGGATTGGGTTAT 1515
300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
1516 CCGAACGGAACCAAGGATGGGCGCGAGTGACCTCGATGATTAATTCGTTGAACGTTGCTTAT 1575
320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
1576 GTGACGATCCAGTGCCTTATCACTAGCAAAAGCGACATATACCTTTACTGCAACG 1635
340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAla 359
1636 GCGGCGCAAGCCATTGAAATCTCCCTGGTATGTCGATGTCCTCGCCCTGCAAGCATCTGCT 1695
360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
1696 TCTGTAACTTGGTTCATGATTTGGATTGGTTCATTACAGCAACCAACGCAAGATAT 1755
380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnValGlu 399
1756 GTCGGGAATGACTTCTCAGCACCATTTTGACAAATACTGGGATGCGCGCAATAACGTAGAA 1815
400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
1816 AATGATTATTATATTCGCCCAAGTGGACATATATACCATTCAGGTGCGCAAGCATATAAT 1875
420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
1876 GTGCGGTTGGACCAACAAACTTCTCGTTGGCAATTGTGAAC 1917
RESULT 10
AR368116 1920 bp DNA linear PAT 12-SEP-2003
LOCUS Sequence 3 from patent US 6376227.
DEFINITION AR368116
ACCESSION AR368116
VERSION AR368116.1 GI:34601777
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1920)
AUTHORS Takaiwa,M., Okuda,M., Saeki,K., Kubota,H., Hitomi,J., Kageyama,Y.,
Shikata,S. and Nomura,M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 3 23-APR-2002;
FEATURES Location/Qualifiers
source 1..1920
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 1,77e-117 Length: 1920
Score: 2002.50 Matches: 383
Percent Similarity: 94.24% Conservative: 26
Best Local Similarity: 88.25% Mismatches: 24
Query Match: 88.57% Indels: 1
DB: 1
Gaps: 1
US-09-985-689A-5 (1-433) x AR368116 (1-1920)
```

QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 616 AATGATGTGGCCAGAGGTAATTGTCAAAGCGGATGTGGCAGCAGCAGCTACGGTTTGTAT 675
QY 21 GlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyValAsnAspSer 40
Db 676 GGACAGGCCAGANTGTCCGAGTGTCCGATACTGGATTGGATACAGAGAGAACGACAGT 735
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCATGAAGCCTCCCGCGTAAATAACAGCACTATATGCACTGGGTCCGACGAAT 795
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 796 AATGCCAATGATACGAACGGTCATGTATCCCATGTGGCAGGTTGGTATTAGGAATGGC 855
QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 856 GCAACGAATAAGGAATGGCACTCAAGCGAATCTGGTTTTCATCTCATCATCGATAGC 915
QY 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn 119
Db 916 AGTGTGGCTTGGAGGTGGCTTCCCAATCTGCAACCTTATTCAGCAAGCAATTCAGT 975
QY 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyValaTyrThra 139
Db 976 GCAGGTGCCAGATTCATACAACTCTCGGGGCGAGCGGTGAATGGGGCTACACGACA 1035
QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAla 159
Db 1036 GATTCAGAAATGTGGATGACTATGTAAGGAAATGATATGACGATTCTTTTCGGGCT 1095
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1096 GGGATGAAGCCGACGCGGTACCATCAGTCACCTGTGTACGGCTAAAGCGCCATA 1155
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
Db 1156 ACAGTCGGCGCAACCGAAACCTGGTCCAGCTTCGTTCCATATGCAGATAATATTAAAC 1215
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyValArgIleLeuProAspVal 219
Db 1216 CAGTTGCACAGTTCTTCCCTCCGCGCGACAAAGATGGCGAATCAAGCCTGATGTC 1275
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 1276 ATGGCCCGCAGGACATACATTTTATCAGCAAGATCTTCTTGTGACCGGATTCCTCCTTC 1335
QY 240 TrrAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
Db 1336 TGGCGGAATCATGACAGCAATATGCTTATATGGTGGNACGCTCCATGGCACACCGATT 1395
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 1396 GTTGGGGGAATGTTGCACAGCTCCGTGAGCATTTTGTGAAAATAGAGGAATCACTCCT 1455
QY 280 LysProSerLeuIleLysAlaIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1456 AAGCCTTCCTATVTGAAAGCAGCTTGTATGAGGTGCTGTGATTTGGATTGGGTAT 1515
QY 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1516 CCGAAGCGAAACCAAGATGGGCGGAGTGACCTGGATAAATCGTTGAACGTTCCCTAT 1575
QY 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
Db 1576 GTGAACGAATCCAGTCCCTATCACTAGCCAAAAGGACATATACCTTTACTGCAACG 1635
QY 340 AlaGlyLysProLeuLysIleSerLeuValTrrThrAspAlaProGlySerThrThrAla 359
Db 1636 GCGGGCAGCAATGAAATCTCCCTGGTATGTCGGATGCGCTGCAAGCACTACTGCT 1695

QY 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1696 TCTGTAAACCTCGTCAATGATTGGATTGGTCATTACGACCAACCGAACAAGATAT 1755
QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrrAspGlyArgAsnAsnValGlu 399
Db 1756 GTCGGGAATGACTTCTCAGCACCATTTGACATAAATCGGATGGCGCGCAATACCTAGAA 1815
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1816 AATGATTATTAAATTCGCCCAAGTGGAAACATATACCATTTGAGTGCACGATATAAT 1875
QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1876 GTCCGGTTGGACACAAAACCTCTCGTTGGCAATTGTGAAC 1917
RESULT 11
LOCUS AR368118 1923 bp DNA linear PAT 12-SEP-2003
DEFINITION Sequence 7 from patent US 6376227.
ACCESSION AR368118
VERSION AR368118.1 GI:34601779
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1923)
AUTHORS Takaiwa,M., Okuda,M., Saeki,K., Kubota,H., Hitomi,J., Kageyama,Y.,
Shikata,S. and Nomura,M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 7 23-APR-2002;
FEATURES Location/Qualifiers
source 1..1923
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores: 2,74e-117 Length: 1923
Pred. No.: 1999:50 Matches: 381
Score: 94.24% Conservative: 28
Percent Similarity: 87.79% Mismatches: 24
Best Local Similarity: 88.43% Indels: 1
Query Match: 6 Gaps: 1
DB: 1
US-09-985-689A-5 (1-433) x AR368118 (1-1923)
QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 619 AATGATGTTCGGCGTGAATTTGTCAAAGCGGATGTGGCTCAGACGACGTACGGGTGTAT 678
QY 21 GlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGACAGGACAGATCGTAGCGGTTGCCGATACAGGCTTGATACAGTCCGCAATGACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 739 TCGATGCATGAAGCCTTCCCGGGGAAAATTTACTGCAATTATATCATTTGGACGCGAAT 798
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 799 AATCCCAATGATACGAATGTGTATGTACGATGTGGTGGCTCCGTTATTTAGGAACGGC 858
QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 859 TCCACTAATAAAGAAAGGCGGCTCAGCGCAATCTAGTCTTCCAATCTATCATGGATAGC 918
QY 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrrAsn 119
Db 919 GGTGGGGACCTTGAGGACTACCTTCGAACTCTCGAACTTATTCAGCAAGCATACAGT 978
QY 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThra 139

Db 979 GCTGTGCGCAGAAATTCATACAACTCTCTGGGACGACGAGTGAATGGGCTTACACA 1038
Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAla 159
Db 1039 GATTCAGAAATGTGGATGATGTCGCGCAAAATGATGACGATCTTTTCGGTGGC 1098
Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaIleVal 179
Db 1099 GGGAAATGAAGACCGAAGCGGGAACCATCATGTCACGAGCAGCTAAATGCAATA 1158
Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnPro 199
Db 1159 ACAGTCGAGAGTACGGAACACTCGCCCAAGCTTTGGGCTTATGGGCAATATCAAC 1218
Qy 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleTyrProAsp 219
Db 1219 CATGTGGCAGCATCTCTCTTCAGTCGAGCGGCAAGGATGACGATCAAAACGGATGC 1278
Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
Db 1279 ATGGCAGCGGAGACCTCATCTATCAGCAAGATCTTCTCTGACCGGATCTCTCTTC 1338
Qy 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
Db 1339 TGGGCGAACCATGACAGATAATATGATACATGCTGGTGAACGCTCCATGGCTAC 1398
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 1399 GTTGTGGAACGTCGACACTCTCTGAGGATTTGTGAAACACAGAGCATCACCA 1458
Qy 280 LysProSerLeuIleLysAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1459 AAGCCTCTCTATTAAAGCGGCACTGATTCGCGGTGACGCTGACATCGGCTTGGCT 1518
Qy 300 ProSerGlyAspGlnGlyTyrGlyValArgValThrLeuAspLysSerLeuAsnValAla 319
Db 1519 CCGACGCTAACCAAGATGGGACGAGTGACATGATGATTAATCCCTGAACGTCCTAT 1578
Qy 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
Db 1579 GTGAACGAGTCCAGTCTCTATCCACCACGCCAAAGCGAGCTACTGCTTACTGCT 1638
Qy 340 AlaGlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrAla 359
Db 1639 GCCGCAAGCTTTGAAATATCTCTGATGATGCTGATGCTGATGCTGATGCTGCT 1698
Qy 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1699 TCCGTACGCTGTCAATGATCTGACCTTCTATACCGCTCCCAATGGCACACAGTAT 1758
Qy 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnValGlu 399
Db 1759 GTAGGAATGACTTTACTTCCCATACATGATGATGATGATGATGATGATGATGAT 1818
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1819 AATGTATTATTATATGACCAACAAAGCGGACGATATCAATGAGTACAGGCTTATA 1878
Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1879 GTACCGGTGGACCAAGCACTTCTCGTTGCAATTTGGAAT 1920

RESULT 12
AX839476
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE

AX839476
Sequence 2 from Patent EP1347044.
AX839476
AX839476.1 GI:39922766
Bacillus sp. KSM-KP43
Bacillus sp. KSM-KP43
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1

AUTHORS
Okuda, M.K., Sato, T.K., Saito, K.K., Sumitomo, N.K., Izawa, Y.K.,
Sasaki, K.K., Kobayashi, T.K. and Nomura, M.K.
TITLE
Alkaline protease
JOURNAL
Patent: EP 1347044-A 2 24-SEP-2003;
Kao Corporation (JP)
FEATURES
Location/Qualifiers
source
1. .1305
/organism="Bacillus sp. KSM-KP43"
/mol_type="unassigned DNA"
/db_xref="taxon:109322"
CDS
1. .1305
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAE85146.1"
/db_xref="GI:39922767"
/translation="NDVARGIVKADVVAOSSYGLYGQGIIVAVDGLDTRNDSSMHE
AFGKITALYALGRTNNANDTNGHGTTHVAGSVLGNSTKMGAPANLVFOSIMDSGG
GLGLSPNLTQLEFQAYSGARLHFNNSGAAYNGAYTTDSRVDDYVKNMTILPAA
GNGPGGCTISAPGTAKNAITVGTENLPPSGSYADNINHVAFSSRGPTKGRKIP
DNVAPGTFILSARSLAPDSSFWANHDSKIATMGSTSMATPIVAGVAQLRHFVKNR
GITPKSLKALIAAGADIGLYPNGNGWRVTLDKSLNVAYNVESSLSSTOKAT
YSTATAGKPKTISLVSDAPASTTASVTLVNDLVLITAPNGTOYVGNDETSPYNDN
WDGRNNVENFVINAPOSYTTIEVOAYNVNVPVPGTFSLAVN"

ORIGIN
Alignment Scores:
Pred. No.: 2e-117 Length: 1305
Score: 1998.50 Matches: 381
Percent Similarity: 94.24% Conservative: 28
Best Local Similarity: 87.79% Mismatches: 24
Query Match: 88.39% Indels: 1
DB: Gaps: 1

US-09-985-689A-5 (1-433) x AX839476 (1-1305)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 1 AATGATGTTCGGGTGGAATGTCAAAGCGGATGTGGCTCAGACGATACGGGTGTAT 60
Qy 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGACAAGACAGATCGTAGCGGTTCGATACAGGCTTGATACAGGTCGAATCAGAGT 120
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCGATGATGAAGCCTTCGCGGGAATTTACTGCAATATATGCAATGGGACGACGAAT 180
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 181 AATGCCAATGATACGAATGCTCATGTACCATGTGGCTGGCTCCGTATTAGGAACGCG 240
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 241 TCCACTAATGAAGAAATGGCGCTCAGGCGAATCTAGTCTTCCATCTATCATGATGAC 300
Qy 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn 119
Db 301 GGTGGGGAATTTGGAGGACTACCTCGAATCTGCAACCTTATTCAGCCAACATACAGT 360
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAla 139
Db 361 GCTGTCGCAAGATTCATCAAACTCTGGGAGCAGCAGTGAATGGGCTTACACAACA 420
Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 421 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGATACGATCCTTTTCGTCGC 480
Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 481 GGGAAATGAAGACCGAAGCGGGAACCATCATGTCGACAGGACGACGCTAAATGCAATA 540
Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199


```
Db 541 ACAGTCGAGGACGACGAAACCTCCGCCACAGCTTTGGTCTTATGCGGCAATATCAAC 600
Qy 200 HisIleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 601 CATGTGGACAGCTTCTCTTCAGTGGACCGCACAAAGGATGGAGGATCAAAACCGGATGTC 660
Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 661 ATGGCACCAGGAACTTCTATCTATCAGCAAGATCTTCTCTGACCGGATCTCTCTTC 720
Qy 240 TrpAlaAsnTyrAsnSerIleTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
Db 721 TGGCGCAACCATGACAGTAATATGATACATAGGTGGAACTCCATCGCTACACCGATC 780
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 781 GTTGCTGGAAACGTGGCAGCTTCTGTAGCACTTTGTGAGAAACAGAGGATCACACA 840
Qy 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 841 AAGCCTTCTCTATTAAGACGCGCACTGATTGCCGTGCGAGTGACATCGGCTTGGCTAC 900
Qy 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 901 CCGAACGTAACCAAGGATGGGACGAGTGACATTGGATAAATCCCTGAACGTTGCCCTAT 960
Qy 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
Db 961 GTGAACGATGCCAGTTCTCTATCCACCGACCAAAAGCGACGTAAGTCTGTTACTGCTACT 1020
Qy 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAla 359
Db 1021 GCGCGCAAGCCTTTGAAATCTCCCTGGTATGCTGTGATGCCCTGCGAGCAACAAGTCT 1080
Qy 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1081 TCGGTAAAGCTTGTCAATGATCTGGACCTTGTCATTCGCTCCAAATGCGCACAGTAT 1140
Qy 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnValGlu 399
Db 1141 GTAGGAAATGACTTTACTTCGCCATACATGATAAATGCGGATGCGCGCAATACGTAGAA 1200
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1201 AATGTATTTATTAATGACCAACAAAGCGGACGTATACAAATGAGGTACAGGCTTATAAC 1260
Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1261 GTACCGGTTGACACACAGACCTTCTCGTTGGCAATTTGTAAT 1302

RESULT 13
AB051423
LOCUS 1923 bp DNA linear BCT 10-MAY-2002
DEFINITION Bacillus sp. KP43 PROF gene for protease, complete cds.
ACCESSION AB051423
VERSION AB051423.2 GI:20521154
KEYWORDS
SOURCE Bacillus sp. KSM-KP43
ORGANISM Bacillus sp. KSM-KP43
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE
1 Itoh, S. and Saeki, K.
AUTHORS new protease
TITLE Unpublished
JOURNAL
REFERENCE 2 (bases 1 to 1923)
AUTHORS Saeki, K.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-2000) Katsuhisa Saeki, KAO CORPORATION;
2606 AKABANE, ICHIKAIMACHI, HAGA, TOCHIGI 321-3486, Japan
(E-mail: 387185@kasanet.kao.co.jp, Tel: 81285687471 (ex. 7471),
Fax: 81285687403)
COMMENT On May 9, 2002 this sequence version replaced gi:14164344.
```

```
FEATURES
source
1..1923
/organism="Bacillus sp. KSM-KP43"
/mol_type="genomic DNA"
/strain="KP43"
/db_xref="taxon:109322"
1..1923
/gene="PROF"
1..1923
/genes="PROF"
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAB5674.2"
/db_xref="GI:20521155"
/translation="MKKKKVLFLSAAAILSTVALSNPSSAGAAFNFLDPKGIQTT
TDKSGKOGTGAALFVSESNVLPKLGKQKLETPVANNKLHI IQFNGLPIBETKQ
QLEKTKAKILDYIPDAIVYIEVGGDVKSATIEHVESVEPLYIRIDPQLFTKGAS
ELYKVALDTKOKNKEVQLEQIAOFAISNDVLYITAKPEVKMNDVARGIVKADV
AQSSYGLYGGQIVAVADTGLDGTQSDSSHEAPRGKITALYALGFTNNANDTGHGT
HVASGVILGGSTNKGMAPQANLVFQSDSGGLGLPSNLQTLFSQAYSAGARHTN
SWGAAYNGAYTDSRNVDYVRKNDMTILFAAGNPGNGGTISAPGTAKNATISVGH
NLSPFSGVADNINHVQFSRRGPTKDRIKPDVWAPGTFILSARSLAPADIGLGPV
DSRYVMGGTSMATPIVAGNVAOLREHFVNKRGITPKSLKALIAADIGLGPV
GNGGWRVLTDKSLNWAYVNSSSLTSOKATYSTFATAGKPLKISLVMSDAPASITA
SVTLVNDLILVITAPNGTQVGNDFTPSYNDNWDGRNVWVFVINAQPSGTYTIEVQA
INVPVGPQFSLAIVN"
ORIGIN
Alignment Scores: 3.17e-117 Length: 1923
Pred. No.: 1998.50 Matches: 381
Score: 94.24% Conservative: 28
Percent Similarity: 87.79% Mismatches: 24
Best Local Similarity: 88.39% Indels: 1
Query Match: 1 Gaps: 1
DB: 1
US-09-985-689A-5 (1-433) x AB051423 (1-1923)
Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 619 AATGATGTGGCGGTGGAATGTCAAGCGGATGTGGCTCAGAGCAGTACGGGTGTAT 678
Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGACAGGACAGATCGTAGCGTTGCCGATACAGGCTTGATACAGGTGCGCAATGACAGT 738
Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 739 TCGATGTCATGAAGCCTTCCCGCGGAAATTAATCTCATTTATATGCTTGGGACGACCAAT 798
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 799 AATGCCAATGATACGAATGGTCATGGTACGCATGTGGCTGGCTCCGTATTAGGAACCGC 858
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 859 TCCACTAATAAGGAATCGCGCTCAGCGCAATCTAGTCTTCCCAATCTATCATGATAGC 918
Qy 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 919 GGTGGGACCTTGGAGGACTACCTTCGATCTGCAACCTTATTTCAGCAAGCATACAGT 978
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAla 139
Db 979 GCTGGTGCAGAAATTCATACAAACTCCTGGGAGCAGCAGTGAATGGGGCTTCACACAA 1038
Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 1039 GATTCCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCTTTTCGCTGCC 1098
Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
```


Db 1099 GGGAAATGAAGGACCGAAGCGGGAACCATCATGTCACCGACAGCTAAATATGCAATA 1158
 QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
 Db 1159 ACAGTCGGAGCTACCGAAACCTCCGCCAAGCTTTGGGTCTTATGGGCAATATCAAC 1218
 QY 200 HisLeuAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
 Db 1219 CATGTGGCACAGTCTCTTTCACGTGGACCGAAGGATGACGGATCAACCGGATGC 1278
 QY 220 ThrAlaProGlyThrPheIleuSerAlaArgSerSerLeuAlaProAspSerPhe 239
 Db 1279 ATGGACCGGGAACCTTCATCTATCACAAGATCTCTCTGACCGGATTCCTCTTC 1338
 QY 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
 Db 1339 TGGGCGAAACCATGACAGTAAATATGCATACATGCGGTGGAACCTCATCGGTACACCGATC 1398
 QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
 Db 1399 GTTGTGGAAAGCTGGACAGCTTCGTGAGCATTTTGTGAAAACAGAGGCATCACACCA 1458
 QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
 Db 1459 AAGCCTTCTCTATTAAAGCGGCACTGATTCGCGGTGCAGCTGACATCGGCTTGGCTAC 1518
 QY 300 ProSerGlyAspGlnGlyTyrGlyValThrLeuAspLysSerLeuAsnValAlaTyr 319
 Db 1519 CCGAAGCTTAACCAAGATGGGACGAGTGACATGTTGTAATATCCCTGAACGTTGCTAT 1578
 QY 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
 Db 1579 GTGAACGAGTCCAGTCTCTATCCACCAAGCAAAAGCGAGTACTCGTTACTGCTACT 1638
 QY 340 AlaGlyLysProLeuLysIleSerLeuValThrThrAspAlaProGlySerThrAla 359
 Db 1639 GCGGCAAGCTTTGAAATCTCCGTGTATGGTCTGATGCCCTGGACACCACTGCT 1698
 QY 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyLysTyr 379
 Db 1699 TCCGTAAACGCTTGTCAATGATCTGGACCTTCTCATATCCGCTCCAAATGGCACACAGTAT 1758
 QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnValGlu 399
 Db 1759 GTAGGAATGACTTTACTTCGCCATACATGATCACTGGATGCCCATTAACGTAGAA 1818
 QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
 Db 1819 AATGTATTATTAAATGCAACCAAGCGGACGATATACAAATTGAGGTACAGGCTTATAAC 1878
 QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
 Db 1879 GTACCGGTGGACACAGACCTTCTCGTTGGCAATTGTGAAT 1920

RESULT 14

AR368117
 LOCUS 1923 bp DNA linear PAT 12-SEP-2003
 DEFINITION Sequence 5 from patent US 6376227.
 ACCESSION AR368117
 VERSION AR368117.1 GI:34601778

KEYWORDS
 SOURCE

ORGANISM

Unclassified.

1 (bases 1 to 1923)

Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y.,

Shikata, S. and Nomura, M.

TITLE Alkaline protease

JOURNAL Patent: US 6376227-A 5 23-APR-2002;

FEATURES Location/Qualifiers

1..1923

/organism="unknown"

/mol_type="genomic DNA"

source

ORIGIN

Alignment Scores:

Pred. No.: 3,17e-117 Length: 1923

Score: 1998.50 Matches: 381

Percent Similarity: 94.24% Conservative: 28

Best Local Similarity: 87.79% Mismatches: 24

Query Match: 88.39% Indels: 1

DB: 6 Gaps: 1

US-09-985-689A-5 (1-433) x AR368117 (1-1923)

QY 1 AsnAspValAlaAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
 Db 619 AATGATGTTCCGTCGTAATTTGCAAGCGGATGTGGCTCAGACGACGTACCGGTGTAT 678
 QY 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGACACGACAGATCGTAGCGGTTCGGATACAGGCTTGATACAGTGCATGACAGT 738
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGATGAGCTTCCCGCGGAAATTTACTGCATTATATGCATTGGACGCGAAT 798
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
 Db 799 AATCCCATGATACGATGCTCATGTACGCAATGTGGCTGGCTCCGTATTAGGAACGCG 858
 QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
 Db 859 TCCACTAATAAGGAATGGCGCCTCAGGCGAATCTAGTCTTCCAACTCATATCATGGATGC 918
 QY 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn 119
 Db 919 GGTGGGAGCTTGAGGACTACCTTCGAATCTCAAACTTATTCAGCAACATACAGT 978
 QY 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrAla 139
 Db 979 GCTGGTCCAGAAATTCATACAACTCTCGGGGACGACGAGTGAATGGGCTTACACACA 1038
 QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
 Db 1039 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCTCTTTTCGTGTC 1098
 QY 160 GlyAsnGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
 Db 1099 GGGAAATGAAGACCGAAGCGGACCATCAGTGCACCGACGACAGCTAAATATGCAATA 1158
 QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
 Db 1159 ACAGTCGGAGCTACGGAAACCTCCGCGCAAGCTTTGGGTCTTATGCGGACAATATCAAC 1218
 QY 200 HisLeuAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
 Db 1219 CATGTGGCACAGTCTCTTTCACGTGGACCGAAGATGAGCGATCAACCGGATGC 1278
 QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
 Db 1279 ATGGACCGGGAACCTTCATCTATCACAAGATCTCTCTGACCGGATTCCTCTTC 1338
 QY 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
 Db 1339 TGGGCGAAACCATGACAGTAAATATGCATACATGCGGTGGAACGTTCCATGCTACCGCATC 1398
 QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
 Db 1399 GTTGTGGAAACGTTGGCACAGCTTCGTGAGCATTTTGTGAAAACAGAGGCATCACACCA 1458
 QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
 Db 1459 AAGCCTTCTCTATTAAAGCGGCACTGATTCGCGGTGCAGCTGACATCGGCTTGGCTAC 1518
 QY 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319

```
Db 1519 CCGAACGTAACCAAGGATGGGACGAGTGCATTGGATTAATCCCTGAACGTTGCCTAT 1578
Qy 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnIysAlaThrTyrSerPheGlnAlaGln 339
Db 1579 GTGAACGAGTCCAGTCTCTATCCACCAAGCAAAAGCGAGCTACTCGTTTACTGCTACT 1638
Qy 340 AlaGlyLysProLeuLysSerLeuValTyrThrAspAlaProGlySerThrAla 359
Db 1639 CCGGCAAGCCTTGAATCTCCCTGATGCTGATGCTGATGCTGCTGCGAGCACAACTGCT 1698
Qy 360 SerTyrThrLeuValAsnAspLeuAspLeuValThrAlaProAsnGlyGlnLysTyr 379
Db 1699 TCCGTAACGCTGTCAATGATCGACCTGTGCTATACCGTCCAAATGGCACACGATAT 1758
Qy 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnValGlu 399
Db 1759 GTAGGAATGACTTTACTTCGCCATCAATGATTAACGCGGCGCAATACGTAGAA 1818
Qy 400 AsnValPheLeuAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1819 AATGATATTATTAATGCACCAAAAGCGGAGCTATACAATGAGGTACAGGCTTATAAC 1878
Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1879 GTACCGTTGGACCAACAGCTTCTCGTTGGCAATGTGAAT 1920

RESULT 15
LOCUS AF268611/c
DEFINITION Uncultured marine group II euryarchaeote EBAC37F11, BAC clone
37F11, partial sequence.
ACCESSION AF268611
VERSION AF268611.1 GI:9664575
KEYWORDS
SOURCE uncultured marine group II euryarchaeote 37F11
ORGANISM Archaea; Euryarchaeota; Marine Group II; environmental samples.
REFERENCE 1 (bases 1 to 60006)
AUTHORS Beja, O., Suzuki, M. T., Koonin, E. V., Aravind, L., Hadd, A.,
Nguyen, L. P., Villacorta, R., Amjadi, M., Garrigues, C.,
Nguyen, L. P., Villacorta, R., Amjadi, M., Garrigues, C.,
Jovanovich, S. B., Feldman, R. A. and DeLong, E. F.
TITLE Construction and analysis of bacterial artificial chromosome
libraries from a marine microbial assemblage
JOURNAL Environ. Microbiol. 2 (5), 516-529 (2000)
MEDLINE 21128653
PUBMED 11233160
REFERENCE 2 (bases 1 to 60006)
AUTHORS Beja, O., Suzuki, M. T., Koonin, E. V., Aravind, L., Hadd, A.,
Nguyen, L. P., Villacorta, R., Amjadi, M., Garrigues, C.,
Jovanovich, S. B., Feldman, R. A. and DeLong, E. F.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-2000) R & D, Monterey Bay Aquarium Research
Institute, P.O. Box 628, Moss Landing, CA 95039-0628, USA
FEATURES
source
1. 60006
/organism="uncultured marine group II euryarchaeote 37F11"
/mol_type="genomic DNA"
/db_xref="taxon:133822"
/clone="BAC 37F11"
/notes="environmental sample"
1484. 3178
/notes="37F11#1"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="S-layer protein"
/protein_id="AAF97179.1"
/db_xref="GI:9664576"
/translations="MBPGQTELIDIFVQSPSAIQVNVKIRVSDSGGQLEEFEI
PVRVSPSPFNASQSWFVQSGEYFLAWMENTGMTSHVTITLENFVGNISGMQ
SASLSPHEARGIPFEIIPPTNWDKQPLELSIVHPQMGVYESTVHSNFSFASP
VLGSGAGSSQAPSLSLNSINPSTVTPDATMNDQVRLILEEGINPETISSTNDAIQ
```

CDS

```
/note="37F11#2"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97180.1"
/db_xref="GI:9664577"
/translations="MNQSNSTSAFYTLDDCCVSNLDETFOALLIRBNHOLWEKLV
NREASTRKQGLDEIDDRNKITWDDATLWETCVRLNRDSDAPIMLGNKNRKFGR
PHTGTSQIYILENISIPQSEGLYSDLIHLEYLTERSEMNTKHKEFLGKAGL
MILGYITPEVKALRSMILGSGMVMVSREPLDGGVREAVRHNLALLMAERRGALIH
ROHE"
complement (3892..4797)
/note="37F11#3"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="SAICAR synthase"
/protein_id="AAF97181.1"
/db_xref="GI:9664578"
/translations="MGEMLYQGVKQVWSTDDPDLLEPFTNQISVFDQIIPSLIPRK
GSLNRTAHWFKLVEQGVCGHNLVFNAPRCLVRKVEVKEGALPRDAEMVFP
LEFIVHYLSGSAWRPFGELTASLVGSQAEYGVKUSEFVEVITFEAFDRNID
REBALQISNITQEFDSIVQAALKYDEIIERAAKNGLIHVGKKEFALGPKRKVLV
DPTGLDEDRWMDAAYSNCECTELSKFVRTHYNSHQELKLARDSGAIDPPIPA
LPESVTEETASLYSMYERLTGQF"
complement (4846..5508)
/note="37F11#4"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="translation initiation factor 6"
/protein_id="AAF97182.1"
/db_xref="GI:9664579"
/translations="MPDILDTHGSDIQIHLSVGNVLFHRELPTPTIEKLEQVVG
LEAPISIGSNLIGALLAGTKGAVADIATESDIDALTAYGDVVMGSGVNTAGNL
LLVNEQGVASPSIPKQGLDIILSEVLGVDIISTTIGGQDVVGSGLVVMGQVLLHPDV
APBEVLLEBILGVPMPVGTSPGSPYVGACASNNAGALAGSETTGPTELNRDIALG
LI"
5597..5745
/product="5S ribosomal RNA"
5816..8611
/product="23S ribosomal RNA"
complement (8888..10192)
/note="37F11#7; contains Zn-ribbon and TPR-like repeats"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97183.1"
/db_xref="GI:9664580"
/translations="MNECAVCGLYLAGSPCCAGCQSLRAQDQSFSDSDALATEVPG
LDDAAQAYDLGIEAPVEQEEBAASLPFGCGESQTNISRLPFGISGPEGIPFEL
SNODSGOEEVHVQHVPEIESTNLEPVALETEQHDSPVYSASVEIPYAPVAPD
LPITBEPVQVPPGPIRLTAIPFVPTVDTSDLPNEMSIESTFGSDQIFABEDVIE
VPSDLEDTVVVEHNGENYQMEKDSVYSEKSLQPFDLHPAQLAVDVGTDASRAL
LESGTISIGQDWRSAARSFQIVARPDGSMNNYGIALLQVAVSMQHSNDPIEVS
NSATQFEAILSLUREAVRTNSQFESYINLSQALLSGREEKALLDLMAGNESQRES
NFTNRAAILAQGRYDEAKRLVPLQADKLASGNLLKJPL"
complement (10248..11969)
/note="37F11#8"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="permease"
/protein_id="AAF97184.1"
```

rRNA

rRNA

CDS

CDS

```
/db_xref="GI:9664581"
/translation="MCGNIFLKRHPAQLIALISLLFVGTQMFIRSHAAGGGDIDL
TVTRQFENQAVYGLIGLILFYALITEVVRHTLAATGLGLALVIALNYSEVAL
SLKAVTIDWETIGLLGMVVGIIISHTGFEPFAVQYKSGGEIWTLVILCSV
TAVLSAFLDNWITMLLTPTVKIAIOVLKPTPILISEVLFNSIGGAATMIGDPPNI
MIGASPSPAISGNKDPNLANASDGVTFNDRTIEMAPGILMTIVPSFMLIRWLKDE
FSGRDRDTELRQVAIKDKLKVSGAILILVILGFLPHVILHAYSWALSALV
MLLNRHELEEELEVEWITLFFAGLFVLSHLSQMGVIDIYGEYVEKAIKFPDD
YRLAAALILWYSASAFIDNPYATMIFIVLSLAFELNPLNPLWALAPGCL
GNGTLIGASANTVAGMEAGYPISFNEFFRAGFPVLLSTFVIFMILVYVVG
EGDALTWKIVLVGISLIGITIAQYSGRAKSPAEALVDVDDFDEIVSTIKSVIKTGS
KGSESE"
complement(12019..12282)
/notes="37Flil149; contains KH domain"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="RNA-binding protein"
/protein_id="AAF97186.1"
/db_xref="GI:9664583"
/translation="MSVPNHI IQAKDTSIPITMRIGKSLTDSVVVELEGOLSSRL
VKVKNRGLYPRDELKTVNIMAEKTNSSVVFARGNVAFWRN"
complement(12279..12614)
/notes="37Flil10; contains Zn-finger; conserved protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97185.1"
/db_xref="GI:9664582"
/translation="MGRSRLNRDKRKHAIWSQTQLLSLLKNPSOFTPEQSRFALH
LVKYSRRHRIIRLSEVEIRMCNTLLRYGENATIRFRNGHKIQTCHSCTSVRRIPY
RHSQKVKA"
complement(12624..13274)
/notes="37Flil11"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="adenylate kinase"
/protein_id="AAF97187.1"
/db_xref="GI:9664584"
/translation="MDEGRQMLNIVLFGPPGAGKGTQQRMMDATGLPQVSTGMDL
RAAKAGTSVLEAKVMDKGVDPDSVIDLRHLDDDKXNVMFDFPRTVPOA
BALSEIVEVSVAIDVPDERIVERIGRYSCTGVVYHDTNPIONGICGCGSTE
KRRADDVEETVLAEAYHTQTSPLADFYRAAGIFHHVDGDRDIDVITDILLAVLG"
complement(13284..16691)
/notes="37Flil12"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="membrane-associated subtilysin-type serine
protease"
/protein_id="AAF97188.1"
/db_xref="GI:9664585"
/translation="MYMOLSGTSQATAVAGGSASLAREYLREVAGINKPSASLIKATL
INGAEDLTPOIPNANEGWGOIDLENSLPSSGVSLSVDFQDDERELOAGFSLIYSD
LDGKNDITLWSDASANAASERLNNLILIIAPDGSYLCNDFSSGISTTG
GSDNNIIRIPAPAGATQGDWVTVHRGGSSORYSIVIAADATLIPKADLTTF
GNSLPSSTELIGLVSISLAWNQGTLAGCPYRIQFEDITSGILYDNRSLLEG
SLSVSFTQSTTGLRSLDNTNNQVPELNAINGDNNIIDLIEITAOGLRV
FQNGQSIPTTSQDRDSAAVTMDVRNETGLSLFPVIAHEGTGRGVPVLSVQSEPD
PTYPILLSPEDSKSVNQGTQVFTISQGGENDTITLNLDDTSASFPQATKRYAR
AGSFIVDYATYQNQPTVSTQRLYIEIGQVDSVDPVPSGTSVSXKPGERTGISGV
RNTGNSAQYTMSCSTSSOWOIMLGNSSSLEFEPLNLIQDLSMDVNVFTPSIANGE
PLAGSDVTCVTSPTSTLNRVTVDAVSELKAFPSDLYDPDGVPGALAPVF
ADTGLVFNHTIQNKGNVLPDPAVTLERGNPKNABIQDEQISLSTVTLGPQOS
GVENRLLVPEPARNSNTYLRVESPQMTNLNLSLVNGENLGNVLSVNSVGTLLA
APVNNDFTFTEIVENSNDLLEWSTSLAPDGSIGSNPPSTVPSVLSQVQLAI
KAPNOTASGFDQMLFVNGTNRFTNABLLVIREVAATSFAGISVDEITAPLFI
PRGDSKQSIITFNGENPLSGLTVEVDGEGNLTDRSSTISPSITDIAIGESIE
VVGKYSTDEDSIDGRMLVILITIDGVVIEFMADTSVSSQSGGIFGILPYLSYP
LVLLALLGVLYGRLKQSSKMDDDGTVELVAPDADTADHGLTREGAOLDISHSVNDI
ASGEVSQBEIAAALQSLDMPVPTKQAVPTGRPPSGLPSNGLPVPVGLPAGLPPAGL
```

```
CDS
PPAKSVPLPVQVTVAGTPPLPPGLPDGWTMEQWNYHGEQVLRMGLEN"
complement(16790..18130)
/notes="37Flil13"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="membrane-associated subtilysin-type serine
protease"
/protein_id="AAF97189.1"
/db_xref="GI:9664586"
Alignment Scores:
Pred. No.: 5,09e-28 Length: 60006
Score: 612.50 Matches: 166
Percent Similarity: 53.30% Conservative: 76
Best Local Similarity: 36.56% Mismatches: 157
Query Match: 27.09% Indels: 56
DB: 1 Gaps: 17
US-09-985-689A-5 (1-433) x AF268611 (1-60006)
QY 13 AlaGlnAsnAsnTyGlyLeuTyGlyGlnGlnValValAlaAlaAAspThrGly 32
Db 17393 GCTCCTCTCTTCTTGAATCTTGATGTTCTCGAGAAACGATTGCGATGATGGATCGC 17324
QY 33 LeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheAArgGlyLeuAla 52
Db 17323 GTCGATATGGAT-----CATCTGACCTCATCGGACGGGTTGCGACA 17282
QY 53 LeuTyAla---LeuGlyArgThrAsnAlaAsnAspProAsn---GlyHisGlyThr 70
Db 17281 ATTATATCTCAGTTCGGTTTAGATAGTTCTCAATTGATAGCAACAGTGGTCTGGGACG 17222
QY 71 HisValAlaGlySerValLeu-----GlyAsnAlaLeuAsnLysGlyMetAla 86
Db 17221 CACGTGGGTATGACTGTGCTTGGCGATGTTACGGGTGATGCGAGTACCGCGGTGTCT 17162
QY 87 ProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeu 106
Db 17161 CAGAAGCAATCTTGTGTATGCTCTTGAGCAGCAGCCACCGGTGTCTTGGAGG 17102
QY 107 ProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIleHisThr 126
Db 17101 CAAGGTTCATCTATGATCTCTCTCTGATCGGAATTAACACAGCAAGTGTCCAGTC 17042
QY 127 AsnSerTrpGlyAlaProValAsn---GlyAlaTyThrAlaAsnSerArgGlnValAsp 145
Db 17041 AATGATGGGGTCTTAATGGAATTAACGGAGCTACACTGCTGA-TCAAGATCGACAGAT 16983
QY 146 GluTyValArgAsnAsnAsp---MetThrValLeuPheAlaAlaGly---AsnGluGly 163
Db 16982 CAATATGTCACCAACAATAATATCTCTCTCTCTTTCGGTGTGGTAAATCGAGGC 16923
QY 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 16922 ACAGGTCTTCAAGATTACTCTCGAGAACAGCAGCAAGATGTTCTTCCATTGGCTCA 16863
QY 184 ThrGluAsnTyArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAlaGln 203
Db 16862 AGCGTTTAAT-----GTTAGC-----GTACGTC 16839
QY 204 PheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaProGly 223
Db 16838 TATTTCACGAGGTCCTCAACTCTCGATGAAGAAATTAACCTGATTAGTCTCCCTCCGCT 16779
QY 224 ThrPheIleLeuSerAlaArgSerLeuAlaProAsp-----GTACGTC 236
Db 16778 GACGGTATCTGTTCTCGACGAGGAGGAAAGCAAGCGTTGTTGGTTTCAGTCTCGCT 16719
QY 237 SerSerPheTrpAlaAsnTyRanSerLysTyAlaTyMetGlyThrSerMetAla 256
Db 16718 ATTGCTACTCACTCAATGCTCGTTCAATGATATATGATATGCAACTCAGCGGAACTTCT 16659
```

QY 257 ThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGly 276
 Db 16658 ACAGCAGTAGCAGGTTGTTTCAGCATCCCTCGCACGTGAATATCTCGGTGAAGTTGCTGGA 16599
 QY 277 IleThr---ProLysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspVal 295
 Db 16598 ATCAACAACCATCTGCAAGCCTCATTAAGGCAACTCTCATCAACGGGGCAGAAAGATTTA 16539
 QY 296 GlyLeu---GlyTyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSer 314
 Db 16538 GGCACACCGGACATTCCTAACCAACACGAGGATGGGGCAAAATGATCTCGAAATAGC 16479
 QY 315 LeuAsnValAla-----TyrValAsnGluAlaThrAlaLeu 326
 Db 16478 CTGAATCTCTTCCTGCTGCGGCTTCTCTGGATGCTTTTCAGGATGATGAGCGTCAACTT 16419
 QY 327 AlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIle 346
 Db 16418 CAGGCAGGATTCTCATTTGATTTATTCCTTTGACCTTGATGGATCAAGGGTATTGACATT 16359
 QY 347 SerLeuValThrAspAlaProGlySerThrThrAlaSerTyrThr-----Leu 363
 Db 16358 ACCTTGGCATGGAGCGATCGAGAAGCCAGTGCAGAAATGCTGCACAATCTGAATCAGTCTA 16299
 QY 364 ValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAsp 383
 Db 16298 TTGAACAATTAGATTGATTCTAATTGCACCCGATGGGTCTTCATCTTGGTAATGAC 16239
 QY 384 PheSerTyrProTyrAsp-----AsnAsnTrpAspGlyArgAsnAsnValGluAsn 400
 Db 16238 TTCTCTCTGGGATTTCACAACAGCGGGTCCCGCTGATAATCTGAACAATATCGAGCGC 16179
 QY 401 ValPheIleAsnAla-----ProGlnSerGlyThrTyrThrIleGluValGlnAla 417
 Db 16178 ATTGGATCCCTCGAGGACGACGACACAGACCGGTGATGGATGTTACAGTTGACAT 16119
 QY 418 TyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
 Db 16118 CGAGGAGGAAGTTCT-----CAGCGCTACAGTATCGTTATT 16083

Search completed: April 4, 2004, 08:08:33
 Job time : 3595.27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 16:37:54 ; Search time 335.908 Seconds
(without alignments)
5476.111 Million cell updates/sec

Title: US-09-985-689A-5
Perfect score: 2261
Sequence: 1 NDVARGIVKADVAQNNYGLY.....EVQAYNVPSQPFSLAIVH 433

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-Q/cor2_1/USPTO spec1/US0985689/runat_31032004_161806_4145/app query.fasta_1.3498
-DB=N_Geneseq_22Jan04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0985689 @CN 1 1 1238 @runat_31032004_161806_4145 -NCPUP=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2245	99.3	1299	2 AAQ27516	AAQ27516 Alkali-pr
2	2044	90.4	3003	2 AAV82382	AAV82382 Bacillus
3	2002.5	88.6	1920	2 AAX37277	AAX37277 Bacillus
4	1999.5	88.4	1923	2 AAX37279	AAX37279 Bacillus
5	1998.5	88.4	1923	2 AAX37278	AAX37278 Bacillus
6	443	19.6	1977	2 AAT85667	AAT85667 Thermococ
7	443	19.6	1977	2 AAX05926	AAX05926 WO9856926
8	422.5	18.7	1236	2 AAX05920	AAX05920 Hyperther

9	422.5	18.7	1566	2	AAT85668	AAT85668 Pyrococcus
10	422.5	18.7	1962	2	AAT85695	AAT85695 Pyrococcus
11	422.5	18.7	1962	2	AAX05929	AAX05929 Hyperther
12	401.5	17.8	1977	2	AAT85669	AAT85669 Protease
13	373.5	16.5	2121	4	ABU54900	ABU54900 T. yonsei
14	345.5	15.3	2539	2	AAT81454	AAT81454 Streptomy
15	345.5	15.3	2809	2	AAT61455	AAT61455 Dnpa-mel
16	341.5	15.1	135638	7	ABX34289	ABX34289 S. atrool
17	334	14.8	1329	6	ABK74643	ABK74643 Bacillus
18	332	13.8	2532	2	AAQ29134	AAQ29134 Encodes R
19	307	13.6	2835	2	AAT08141	AAT08141 Hyperther
20	307	13.6	4765	2	AAT08132	AAT08132 Protease
21	307	13.6	4765	2	AAT85670	AAT85670 Pyrococcus
22	306.5	13.6	1859	2	AAT85677	AAT85677 Thermococ
23	306.5	13.6	1859	2	AAT85677	AAT85677 Thermococ
24	298.5	13.2	3743	9	ADD24901	ADD24901 DNA encod
25	295	13.0	1560	7	ABZ37569	ABZ37569 Streptomy
26	285	13.0	59816	7	ABZ37516	ABZ37516 Streptomy
27	285	13.0	59816	7	ABZ37515	ABZ37515 Hyperther
28	283	12.5	898	2	AAT08131	AAT08131 Hyperther
29	282	12.5	564	2	AAT08134	AAT08134 DNA seque
30	282	12.5	564	2	AAT85676	AAT85676 Thermococ
31	281.5	12.5	546	4	ABL53453	ABL53453 T. yonsei
32	276.5	12.2	2166	6	ABU40498	ABU40498 DNA const
33	276.5	12.2	3788	9	ADD24905	ADD24905 DNA encod
34	273	12.1	807	2	AAQ11835	AAQ11835 Carbonyl
35	270	11.9	810	7	ABZ21470	ABZ21470 Mature Ba
36	269	11.9	866	2	AAQ30104	AAQ30104 Synthetic
37	269	11.9	866	2	AAQ30146	AAQ30146 Synthetic
38	269	11.9	866	2	AAQ31697	AAQ31697 Subtilisi
39	269	11.9	866	2	AAQ26410	AAQ26410 Subtilisi
40	269	11.9	866	2	AAQ26410	AAQ26410 Subtilisi
41	269	11.9	1140	2	AAT97502	AAT97502 Bacillus
42	269	11.9	1140	2	AAT93806	AAT93806 Bacillus
43	269	11.9	1140	2	AAV69598	AAV69598 B. lentus
44	269	11.9	1140	3	AAZ59988	AAZ59988 Bacillus
45	269	11.9	1140	3	AAA59767	AAA59767 Subtilisi

ALIGNMENTS

RESULT 1
AAQ27516
ID AAQ27516 standard; DNA; 1299 BP.

AC AAQ27516;

DT 05-FEB-1993 (first entry)

DE Alkali-protease Ya enzyme gene.

KW Alkali resistance; surface active agent resistance; detergency improver;
KW ss.

OS Bacillus sp. Y.

PH Key Location/Qualifiers

FT CDS 1..1299

FT /*tag= a

PN JP04197182-A.

PD 16-JUL-1992.

PF 28-NOV-1990; 90JP-00327110.

PR 28-NOV-1990; 90JP-00327110.

PA (LIOY) LION CORP.

XX WPI; 1992-288440/35.

DR P-PSDB; AAR26274.

PT DNA coding alkali-protease Ya enzyme - has good alkali and surfactant
 resistance and improves detergency.

XX Claim 3; Page 2; 17pp; Japanese.

CC The sequence is that of the alkali-protease Ya enzyme gene which can be
 used in the recombinant production of Ya enzyme. Ya enzyme is excellent
 in alkali resistance and surface active agent resistance and improves
 detergency

SQ Sequence 1299 BP; 401 A; 246 C; 297 G; 355 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.4e-168 Length: 1299
 Score: 2245.00 Matches: 429
 Percent Similarity: 99.77% Conservative: 3
 Best Local Similarity: 99.08% Mismatches: 1
 Query Match: 99.29% Indels: 0
 DB: 2 Gaps: 0

US-09-985-689A-5 (1-433) x AAQ27516 (1-1299)

QY 1 AsnAspValAlaArgGlyIleValIleAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
 Db 1 AATGATGTAGCAAGAGGATAGTAAAGCTGATGTCACAAAACAATACGGATTATAT 60
 QY 21 GlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 61 GGACNAGTCACTAGTTCAGTAGCGGACACAGGCTTAGTACAGGTCTGACGATAGT 120
 QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGTCATGAAGCATTCGCGGGAATAATCACAGCTCTTTACGCGTTAGGAACAATAAT 180
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyValAla 80
 Db 181 AATGCGAGTATCCGATGGCATGGCATGCGACATGATGACAGGTTCTGTACTTGTATGCT 240
 QY 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
 Db 241 TTAATAAAGGAATGGCTCCCGAGCTAACTTAGTCTTCCAAATCTATTATGGATAGCAGC 300
 QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAla 120
 Db 301 GGAGATTAGTGGCTTACCATCGAATTAATACGTTATTTAGTCAAGCTTCGAATGCT 360
 QY 121 GlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
 Db 361 GGAGCAAGAATTTCATTAATCTTGGGGAGCCCGATTAATGGAGCGTACACTGCTAAC 420
 QY 141 SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
 Db 421 TCGAGCAAGTGGATGAGTATGTTCCGAATAATCATATGACGGTACTTTTTCAGCTGGT 480
 QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 Db 481 AATGAAGTCTCTAATTCAGGAACAATTAGTCTCCAGGTACAGCGAAAAATGCTATTACG 540
 QY 181 ValGlyValThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
 Db 541 GTCCGCGCAACGGAATACTATCGCCCAAGCTTCGGTTCGATAGCAGATAACCCAAATCAT 600
 QY 201 IleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 Db 601 ATTGCACAATTTTCATCAGAGGAGCTACGAGGAGATGGACGAATTAAGCCCTGACGTAACA 660
 QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr 240
 Db 661 GCTCTCGGACATTTATTTATCAGCACGTTCTTCCCTAGTCCAGACTCTTCGTTTGG 720
 QY 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
 Db 721 GCGAATTATAACAGTAAATACGCGTATATGGGCGGTACCTCCATGGCGACACCTATTGTT 780

QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
 Db 781 CGAGGAATGTCCGCAATTAGTGAGCATTTTATAAAAAATAGAGTATTACTCTCTAG 840
 QY 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
 Db 841 CCTCTCTTATAAAGCTGCTATTCGCTGTGTACTGATGTTGGTTTAGGATATCCT 900
 QY 301 SerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
 Db 901 ASTGTTGACCAAGCTGGGGCGGTGTTACTCTAGATTAATCGTTAATGTAGCGTATGTC 960
 QY 321 AsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAla 340
 Db 961 AATGAAGCAACTGCTCATTTAGCCACAGACAAAAGCAAGTATTCGTTCCAGGCAACAGC 1020
 QY 341 GlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAlaSer 360
 Db 1021 GGTAAACCTTAAAAATCTCGTTAGTATGACAGATGCTCTCTGGAAGTACAACTGCACT 1080
 QY 361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
 Db 1081 TATACACTAGTTAATGATTAGTATCTAGTATCTAGTCTCCGAATGGACAAAATATGTA 1140
 QY 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnValGluAsn 400
 Db 1141 GGAATGATTTTATGTTATCTTATGATAATAACTGGATGGTCCGCAACAATGTTGAGAAC 1200
 QY 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValAlaIleThrVal 420
 Db 1201 GTATTATAAACCTCCGCAATCTGGAAGTATATAATTGAGGTTCAGGCGTATTAATGTA 1260
 QY 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
 Db 1261 CCATCTGCCCCACAGCGTTTCTCCTAGTATCGTACAT 1299

RESULT 2
 AAV82382
 ID AAV82382 standard; DNA; 3003 BP.
 XX
 AC AAV82382;
 XX
 DT 12-APR-1999 (first entry)
 XX
 DE Bacillus JP170 protease gene.
 XX
 KW Protease; detergent; surfactant; leather processing; debittering;
 KW flavour; ss.
 XX
 OS Bacillus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 846..2771
 FT /*tag= a
 FT sig_peptide 846..944
 FT /*tag= b
 FT mat_peptide 1470..2768
 FT /*tag= c
 XX
 FT WO9856927-A2.
 PN
 XX 17-DEC-1998.
 PD
 XX 09-JUN-1998; 98WO-US012005.
 PF
 XX 12-JUN-1997; 97US-00873479.
 PR (NOVO) NOVO NORDISK BIOTECH INC.
 XX
 PA Sloma A, Christianson L;
 PI WPI; 1999-080908/07.
 XX
 DR

DR P-PSDB; AAW89547.
 XX Novel protease from *Bacillus subtilis* LC20 - useful in laundry and
 PT dishwashing detergents and for leather processing.
 XX
 XX Claim 11; Page 52-53; 77pp; English.
 XX
 CC This nucleotide sequence encodes a novel protease (see AAW82382) of
 CC *Bacillus* sp. JPI170 (NCIB 12513). The sequence in plasmid p170BAN is
 CC contained in *Bacillus subtilis* LC20 NRRL B-21680. The protease gene was
 CC isolated from chromosomal DNA of JPI170 following preparation of probes
 CC based on protease N-terminal and internal peptides (see AAW89549-50),
 CC screening of chromosomal libraries, isolation of the 3' end of the gene
 CC by inverse PCR (see AAW82410-11), reconstruction of the 5' and 3' ends and
 CC PCR amplification (see AAW82412-16). Claimed recombinant host cells can
 CC be used in a method for producing the protease. The protease is used in
 CC laundry and dishwashing detergents, for institutional and industrial
 CC cleaning, and for leather processing, as well as for debittering and
 CC enhancing the degree of hydrolysis of protein hydrolysates, for flavour
 CC development through hydrolysis of proteins, degradation of undesired
 CC peptides and in enzymatic synthesis of peptides. It has enhanced
 CC stability towards oxidation under alkaline conditions, e.g. towards
 CC bleaching agents of the peroxy type. The invention also provides mutant
 CC cells in which the protease activity is diminished. Such cells can be
 CC used for the production of heterologous recombinant proteins
 XX
 SQ Sequence 3003 BP; 982 A; 504 C; 645 G; 872 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3,2e-152 Length: 3003
 Score: 2044.00 Matches: 386
 Percent Similarity: 94.69% Conservative: 24
 Best Local Similarity: 89.15% Mismatches: 23
 Query Match: 90.40% Indels: 0
 DB: Gaps: 0

US-09-985-689a-5 (1-433) x AAW82382 (1-3003)

QY 1 AsnAspValAlaArgGlyValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
 DB 1470 AATGACGTGGCGCGTGGCATTCGAAAGCAGACGTCGACAAATAAATCTTGGCTTATAT 1529
 QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyValAsnAspSer 40
 DB 1530 GGACAAAGGACAGATTGTAGCATTTGCTGATCTGGCGTGTATACAGGAAGAAATGACAGT 1589
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 DB 1590 TCGATGCATGAAGCATTCGCGGTGAAGATTACCGCACATATATGCACTGGGAGAACGAAT 1649
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyValAla 80
 DB 1650 AAGCCCAATGATCCAAATGGACATGGACCCATGTTCGTGGATCTGTGTAGGAATGCT 1709
 QY 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
 DB 1710 ACAATAAAGGATGGCAGCGCAAGCAATCTAGCTTTTCAATCTATTATGATAGTGGT 1769
 QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaThrAsnAla 120
 DB 1770 GGAGGGCTGGGAGGACTACTGCTTAATCTACAAACATTTATTCAGTCAAGCATATAGTGT 1829
 QY 121 GlyValaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlnValAlaTyrThrAla 140
 DB 1830 GGAGCAGAGATTCTATACGAATTCATCGGGGGCTCCAGTAAACGGTCCCTATACGACAGAC 1889
 QY 141 SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
 DB 1890 TCTCGAAATGTTGATGATTATGTGAGAAAAAATGATATGACGATTTCTTTTCGGCCGGA 1949
 QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 DB 1950 AATGAGGGACCGGTACGGGTACATCAGTCGACACGACAGACAAATAATGGATTACA 2009

QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
 DB 2010 GTTGGGGACCGAAACCTAGCTTCAGCTTCGGATCTTATCGGATAAATATTAAACCAT 2069
 QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLeuProAspValThr 220
 DB 2070 GTTGTCAATCTCTTCACGAGTCTACTAGATGGACGCTATTAAAGCCGACGTCATG 2129
 QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr 240
 DB 2130 GCACCAAGGTACGATCTCTCTCTAGATCATCATTTAGCTCCAGATCTCTCATCTTGG 2189
 QY 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
 DB 2190 GCAACCATGATAGTAAATATGCTTACATGGGTGCTACTTCTATGGCTACTCAATTGTA 2249
 QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
 DB 2250 GCAGGTAAATGTTGCCAATTAAGGGAGCATTTTGTGAAAAATAGAGGGGTACTCTTAAG 2309
 QY 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
 DB 2310 CTTCCCTTTTAAAGCTCTTTAATTCAGGTCTGCGGATGTTGGACTTGGCTTTCCA 2369
 QY 301 SerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
 DB 2370 AATGTAAACCAAGGATGGGAAAGATACGTTAGATAAATCCCTAAATGTCGATTTTIG 2429
 QY 321 AsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAla 340
 DB 2430 AATGAACGAGCCCTTTATCAACAGTCAAAAGCAACATATTCTGTTTACGGCTCAAGCT 2489
 QY 341 GlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAlaSer 360
 DB 2490 GGTAAACCCCTTAAAAATATCATCTTGTTCGATGCCAGGATGACGACGAGGATCA 2549
 QY 361 TyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
 DB 2550 CTAACTTTAGTGAATGATTTAGACTTAGTAATCACTGCACCAATGAACCTAAATACGTC 2609
 QY 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnAsnValGluAsn 400
 DB 2610 GGAAATGACTTTACAGCACCGTATGATAACAATTTGGATGGCAGAAACACGTTGGAAT 2669
 QY 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
 DB 2670 GTGTTTATCAATGCTCTCAAGCGGAACGTATACAGTCGAAGTCGAGCTTACATGTA 2729
 QY 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
 DB 2730 CCAGTAAGTCCGCAACCTTTCTTTTAGCGATTGTACAT 2788

RESULT 3
 AAX37277 standard; DNA; 1920 BP.
 ID AAX37277
 XX AC AAX37277;
 XX DT 20-MAR-2003 (revised)
 XX DT 21-JUL-1999 (first entry)
 XX DE Bacillus alkaline protease encoding DNA.
 XX KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 XX KW washing composition; oxidising agent; ss.
 XX OS Bacillus sp.
 XX PN WO9918218-A1.
 XX PD 15-APR-1999.

PF 07-OCT-1998; 98WO-JP004528.
 XX 07-OCT-1997; 97JP-00274570.
 PR (KAOS) KAO CORP.
 XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX WPI: 1999-287736/27.
 DR P-PSDB; AAY17087, AAY17089.
 XX Alkali protease from *Bacillus* used in washing powders.
 PT Disclosure; Page 53-58; 71pp; Japanese.
 XX The invention relates to alkaline proteases produced by strains of
 CC *Bacillus*. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-12; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 1920 BP; 576 A; 411 C; 473 G; 460 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3,6e-149 Length: 1920
 Score: 2002.50 Matches: 383
 Percent Similarity: 94.24% Conservative: 26
 Best Local Similarity: 88.25% Mismatches: 24
 Query Match: 88.57% Indels: 1
 DB: 2 Gaps: 1

US-09-985-689A-5 (1-433) x AAX37277 (1-1920)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
 Db 616 AATGATGTGCCAGAGGTATTGTCAAAGCGGATGTGGCACAGAGCACTACGCTTTGTAT 675
 QY 21 GlyGlnGlyValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 676 GGAACAAGCCAGATTGTCCGAGTGTCCGATCTGGATTGGATTGATACAGAAACACGACGT 735
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 736 TCGATGATGAGGCTTCGCGGTAAATAACAGCACTATATGCACTGGGTGCGACCAAT 795
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
 Db 796 AATCGGAATGATACGAACGGTCTATGTPACCCATGTGGCAGGTTCCGGATTAGGAATGCG 855
 QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
 Db 856 GCAACGATTAAGAAATGACCTCAAGCGAATCTGGTTTTTCAATCCATCATGATAGC 915
 QY 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
 Db 916 AGTGGTGGCTTGGAGGCTTGCCTTCCAAATCTGCAACCTTATTCAGCCAAAGCATTCAGT 975
 QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValAlaProValAsnGlyAlaTyrThrAla 139
 Db 976 GCAGTGGCCAAATTCATCAAACTCTGGGGGCGACGGTGAATGGGGCTACACGACA 1035
 QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159

Db 1036 GATTCAGAAATGTGATGACTATGTAAGGAAAAATGATATGACGATTCCTTTTCGGCGCT 1095
 QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
 Db 1096 GGGATGAAGCCGAAACGGCGGTACCATCACTAGTCACCTGGTACGGCTAAAAACCCATA 1155
 QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
 Db 1156 ACAGTCGGCGCAACCGAAACCTGGTCCAGCTTCGGTTCCTATGCAGATAATATTAAAC 1215
 QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
 Db 1216 CACGTTGCACAGTCTCTTCCCGTCCCGCAGCAAAAGATGGCGCAATCAAGCTCATGTC 1275
 QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
 Db 1276 ATGGCGCCAGGACATACATTTTATCAGCAAGATCTTCTCTTGCACCCGATTCCTCTTC 1335
 QY 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
 Db 1336 TGGCGCAATCATCAGCAAAATATGCTATATGGGTGGAACTCCATGTCACACCGAT 1395
 QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
 Db 1396 GTTCCGGGGAATTTGCACAGCTCCCGTGGCATTTTGTGAAAATATAGAGATCATCTCT 1455
 QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
 Db 1456 AAGCCTTCCTTATTTGAAGCAGCTTTGATTCGAGGTGCTGCTGATTTGGATTGGTTAT 1515
 QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
 Db 1516 CCGAACGGAACCAAGGATGGGCGCGAGTGACCTGGATAATTCGTTGAACGTTGGCTAT 1575
 QY 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
 Db 1576 GTGACGAAATCCAGTGCCCTATCACTAGCCAAAAGCGACATATACCTTTACTGCAACG 1635
 QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAla 359
 Db 1636 GCGGCAAGCCATTTGAAAATCTCCCTGTATGTGCGATGCCCTCCCTCAAGCATCTAGCT 1695
 QY 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
 Db 1696 TCTGTAAACCTGGTCAATGATTTGGATTGGTTCATTACAGCAACCAACGAAACAGATAT 1755
 QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
 Db 1756 GTCGGAATGACTTCTCAGCACCATTTGACAATAACTGGGATGGCGCAATAACGTAGAA 1815
 QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
 Db 1816 AATGTATTTTATTAATTCGCCCAAGTGGACATATATACCATTTGAGGTGCAAGCATATAAT 1875
 QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
 Db 1876 GTGCGGTTGGACCAAAAACCTTCTCGTTGGCAATTTGTGAAC 1917

RESULT 4
 AAX37279
 ID AAX37279 standard; DNA; 1923 BP.
 XX
 AC AAX37279;
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE *Bacillus* alkaline protease encoding DNA.
 XX
 KW Alkaline protease; *Bacillus*; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent; ss.
 XX
 OS *Bacillus* sp.

XX WO9918218-A1.
 XX PD 15-APR-1999.
 XX PF 07-OCT-1998; 98WO-JP004528.
 XX PR 07-OCT-1997; 97JP-00274570.
 XX PA (KAO) KAO CORP.
 XX PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 XX PI Shikata S, Nomura M;
 XX DR WPI; 1999-287736/27.
 XX PR P-PSDB; AAY17091.
 XX PI Alkali protease from *Bacillus* used in washing powders.
 XX PS Disclosure; Page 63-68; 71pp; Japanese.
 CC The invention relates to alkaline proteases produced by strains of
 CC *Bacillus*. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX SQ Sequence 1923 BP; 578 A; 417 C; 474 G; 454 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 6,246-149 Length: 1923
 Score: 1999.50 Matches: 381
 Percent Similarity: 94.24% Conservative: 28
 Best Local Similarity: 87.79% Mismatches: 24
 Query Match: 88.43% Indels: 1
 Ds: 2 Gaps: 1

US-09-985-689A-5 (1-433) x AAX37279 (1-1923)

QY 1 AsnAspValAlaArgGlyVileValIysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
 Db 619 AATGATGTTGCGGTGGAAATTGTCAAAGCGGATGTGGCTCAGAGCAGCTACGGGTTGTAT 678
 QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGCAAGACAGATCGTACGGGTTCCCATACAGGCTTGATACAGGTGCGCAATGACAGT 738
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGCATGAGCCTTCGCGGGAAATTAATCTGATTAATGATTTGGGACGACGAT 798
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
 Db 799 AATGCCAATGATPACGAATGGTTCATGATACGATGTGGCTGGCTCCGATTATAGGAACCGC 858
 QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
 Db 859 TCCACTAATAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCAACTCATCATGGATAGC 918
 QY 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
 Db 919 GGTGGGGACTTGGAGGACTACCTTCGAATCTGCAACCTTATTTCAGCAAGCATACAGT 978
 QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValaTyrThrAla 139

Db 979 GCTGTGTCAGAAATTCATACAACTCCTGGGAGCAGCAGTGAATGGGGCTTACACAACA 1038
 QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
 Db 1039 GATTCAGAAATGTGGATGACTATGTGGCAAAATGATATGACGATCCTTTTCTGCTGCC 1098
 QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
 Db 1099 GGAATGAAGACCGAACCGCGGAACCATCAGTCACCCAGGCACAGCTAAAAATGCAATA 1158
 QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
 Db 1159 ACAGTCGGAGCTACGGAACCTCCGCCCAAGCTTTGGTCTTATGCGGACAAATATCAAC 1218
 QY 200 HisIleAlaGlnPheSerArgGlyAlaThrArgAspGlyValGlyIleLysProAspVal 219
 Db 1219 CATGTGGCACAGTTCTCTTCAGTGGACCGACAAAGGATGGACGATCAAAACCGGATGTC 1278
 QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
 Db 1279 ATGGCACCGGAAACGTTTATATCAGCAAGATCTTCTTTGCACCGGATTCCTCCTTC 1338
 QY 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
 Db 1339 TGGCGAACCATGACAGTAATATGATACATGCGGTGGACGTCCTCATGGCTACACGATC 1398
 QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
 Db 1399 GTTGTCTGGAACCGTGGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCATCACCCA 1458
 QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
 Db 1459 AGCCTTCTCTATTAAAGCGGACATGATGCGGTGCGAGCTGACATCGGCCCTTGGCTAC 1518
 QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
 Db 1519 CCGAACCGTAAACCAAGGATGGGACGAGTGACATTTGGATAAAATCCCTGAACGTTGCCTAT 1578
 QY 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
 Db 1579 GTGAACGATGCCAGTTCTCTATCCACCGCAAAAGCGACGATCTCGTTTACTGCTACT 1638
 QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAla 359
 Db 1639 GCGGCAAGCCTTGAATAATCTCCCTGTGTATGCTGTGATGCGGCTGCGAGCAACTGCT 1698
 QY 360 SerTyrThrIleuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
 Db 1699 TCGTAAACGCTTGTCAATGATCTGGACCTTGTCTATTACCGCTCCAAATGGCACAGTAT 1758
 QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
 Db 1759 GTAGGAATGACTTACTTCCGCATACATGATGATGCGGATGGCGCAATACGTAAGA 1818
 QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
 Db 1819 AATGTATTTAATTAATGACCAACAAAGCGGACGATACAAATGGAAGTACAGGCTTATAAC 1878
 QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
 Db 1879 GTACCGGTTGGACCAACAGAACTTCTCGTTGGCAATTGGAAT 1920

RESULT 5
 AAX37278
 ID AAX37278 standard; DNA; 1923 BP.
 XX AC AAX37278;
 XX DT 20-MAR-2003 (revised)
 XX DT 21-JUL-1999 (first entry)
 DE *Bacillus* alkaline protease encoding DNA.

XX Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent; ss.
 XX Bacillus sp.
 OS WO9918218-A1.
 PN 15-APR-1999.
 XX 07-OCT-1998; 98WO-JP004528.
 XX 07-OCT-1997; 97JP-00274570.
 XX (KAOS) KAO CORP.
 XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX WPI; 1999-287736/27.
 DR P-PSDB; AAY17088, AAY17090.
 XX Alkali protease from Bacillus used in washing powders.
 PT Disclosure; Page 58-63; 71pp; Japanese.
 XX The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX Sequence 1923 BP; 576 A; 418 C; 475 G; 454 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 7 49e-149 Length: 1923
 Score: 1998.50 Matches: 381
 Percent Similarity: 94.24% Conservative: 28
 Best Local Similarity: 87.79% Mismatches: 24
 Query Match: 88.39% Indels: 1
 DB: 2 Gaps: 1

US-09-985-689A-5 (1-433) x AAX37278 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAspThrGlyLeuTyr 20
 Db 619 AATGATGTTGGCGTGGAAATGTCAAAGCGGATGGCTCAGACGACGTACGGGTGTAT 678

Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGACAGGACAGATCGTACCGGTTGCCGATACAGGGTTGNTACAGGTCCGATGACAGT 738

Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGATGAAGCCCTCCGCGGGAAATTAAGTCAATATATGTCATTTGGGACGCGCAAT 798

Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
 Db 799 AATGCCAATGATACGAATGGTTCATGTCAGCATGTCGCTGGCTCCGTTATAGGAACGGC 858

Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
 Db 859 TCCACTAATAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCAAATCTATCATGGATAGC 918

Qy 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
 Db 919 GGTGGGGGACTTGGAGGACTTACTTTCGAATCTGCAAACTTATTAGCCAGCATACAGT 978

Qy 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAla 139
 Db 979 GCTGGTGCCAGAAATTCATCAAACTCTCTGGGAGCAGCAGTGAATGGGGCTTACACACA 1038

Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAla 159
 Db 1039 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCCTTTTGGCTGC 1098

Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
 Db 1099 GGGAAATGAAGACCGAAACGGCGGAACCATCAGTGCACCGAGGACACCTTAAATGCAATA 1158

Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
 Db 1159 ACAGTCGAGCTACGGAAACCTCCGCCCAAGCTTTGGTCTTTATGCGGACAATATCAAC 1218

Qy 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
 Db 1219 CATGTGCACAGTTCTCTTCACGTGACCGACCAAGATGTCGATCAAAACCGGATGTC 1278

Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
 Db 1279 ATGGACCGGGAACGTTTCATCTATCAGCAAGATCTTCTTCACCGGATTCCTCCTTC 1338

Qy 240 TrpAlaThrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
 Db 1339 TGGCGGAACCATGACAGTAATAATGATGATGATGCGTGGAAACGTCCTGATGGTACACCATC 1398

Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
 Db 1399 GTTCTGTAACGTTGGGACAGCTTCGTCAGCATTTTGTGAAACACAGAGCATCACACCA 1458

Qy 280 LysProSerLeuIleLysAlaLeuAlaGlyAlaThrAspValGlyLeuGlyTyr 299
 Db 1459 AAGCCTTCTCTATTAAAGCGGCACACTGATTCGCGGTGACGTGACATCGGCTTGGGTAC 1518

Qy 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
 Db 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATTTGATTAATCCCTGAACGTTGCCTAT 1578

Qy 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
 Db 1579 GTGAACGAGTCCAGTTCTCTATCCACCGCCAAAGCGACGCTACTCGTTTACTGCTACT 1638

Qy 340 AlaGlyLysProLeuLysIleSerLeuValThrPheAspAlaProGlySerThrAla 359
 Db 1639 GCCGGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCCCTCGGAGCAGCACTGCT 1698

Qy 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
 Db 1699 TCCGTAAACGTTGTCAATGATCTGACCTTGTCTTACCGCTCCAAATGGCACAGTAT 1758

Qy 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnThrAspGlyArgAsnValGlu 399
 Db 1759 GTAGGAATGACTTTTACTTCGCCATACAACTGATTAACCTGGATGGCCGCAATACTAGNA 1818

Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
 Db 1819 AATGATTATTATTAATGACCCACCAAGCGGACGTATACAAATTGAGGTACAGGCTATAAC 1878

Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
 Db 1879 GTACCGGTGGACCAACAGACCTTCTCGTTGGCAATTTGTAAT 1920

RESULT 6
 AAT85667
 ID AAT85667 standard; DNA; 1977 BP.
 XX
 AC AAT85667;

XX 17-OCT-2003 (revised)
 DT 20-APR-1998 (first entry)
 XX Thermococcus protease coding sequence.
 DE
 KW Protease; research reagent; thermal stability; thermococcus celer; ss.
 XX
 XX Thermococcus celer; DSM-2476.
 OS
 XX WO9721823-A1.
 PN 19-JUN-1997.
 XX
 XX 07-NOV-1996; 96WO-JP003253.
 PF
 XX 12-DEC-1995; 95JP-00323285.
 PR
 XX (TAKI) TAKARA SHUZO CO LTD.
 PA
 XX Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
 PI Tsunawasa S, Kato I;
 PI
 XX WPI; 1997-332794/30.
 DR P-PSDB; AAW24121.
 XX
 XX Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT industrially and as research reagents.
 XX
 XX Claim 3; Page 86-87; 159pp; Japanese.
 PS
 XX This sequence represents the coding sequence for the protease from
 CC Thermococcus celer DSM-2476. This sequence encodes a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries. (Updated on 17-
 CC OCT-2003 to standardise OS field)
 XX
 XX Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 1,22e-25 Length: 1977
 Score: 443.00 Matches: 137
 Percent Similarity: 46.03% Conservative: 66
 Best Local Similarity: 31.07% Mismatches: 170
 Query Match: 19.59% Indels: 68
 DB: 2 Gaps: 16

US-09-985-689A-5 (1-433) x AAT85667 (1-1977)

QY 8 ValAlaAspValAlaGlnAsnTyrGlyLeuTyrGlyGlnGlnValValAla 27
 DB 433 ATAGGGCGCGATACCGTCTGGAACTCCCTCGGTACGACGAGCGGTGTGGTTGCC 492
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
 DB 493 ATCGTCGATACGGGTATAGACGCGAAC-----CACCCCGCATCTGAAG 534
 QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
 DB 535 GGCAGGTCTAGCTGTACGACGCGGTCAACGGCAGGTGACCCCTACGATGACCAG 594
 QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
 DB 595 GGACACGGAACCCACGTTGCGGGTATCGTTGCGGAACCGGACGCGTAACTCCCAAGTAC 654
 QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 100
 DB 655 ATAGGCGTCCGCGCGGCGGAGCTGTCGGCGTCAAGGTTCTCGGTCGCGGACGTTG 714
 QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120

Db 715 GGAAGCGTCTCCACCATCATCGCGGTGTTGACTGGTCTGCAGAACCAAGACAGTAC 774
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
 DB 775 GGGATTAAGGGTCATCAACCTCTCCCTCGGTCTCTCCAGAGCTCCGACGGAACCGACTCC 834
 QY 141 SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
 DB 835 CTCAGTCAGGCGGTCAACCAACCGCTGGGACGCGGTATAGTAGTCTGCTGCGCGCGGC 894
 QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 DB 895 AACAGCGGCGCGAACACCTACCGTGGCTCACCGCGCGCGGAGCAAGGTCAATACC 954
 QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
 DB 955 GTCGGTGCA-----GTTACAGCAACGACAAC 981
 QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 DB 982 ATGCCAGCTTCTCCAGCGGACCGACCGCGGACGGAAGGTCAAGCCGGAAGTCGTC 1041
 QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
 DB 1042 GCGCGCGGTGACATCATAGCCCGCGCGCAGC-----GGAACCAAGCATGGGC 1092
 QY 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
 DB 1093 ACCCGATAAACCACTACTACCAAGGCTCTGGAAACCAAGATGCCACCGCGACGTT 1152
 QY 261 AlaGlyAsnValAlaGlnLeuArgGlyHisPheIleLysAsnArgGlyIleThrPro--- 279
 DB 1153 TCGGCGCTTGGCGGTCTATCTCCAG-----GCCACCGCGAGTGGACCCCGGAC 1203
 QY 280 -----LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly--- 296
 DB 1204 AGGTGAAGACCGCGCTCATCGAGACCGCGCATAGTGTCCCGCCAGGATAGCGGAC 1263
 QY 297 LeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsn 316
 DB 1264 ATCGCTAC-----GGTGGGTAGGTGAACGTCTACAGGCG----- 1302
 QY 317 ValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGln-----Lys 331
 DB 1303 ATCAAGTACGACGACTACGCCAAGCTCCCTTCCCGGCTCCGTGCGCGCAAGGGAAGC 1362
 QY 332 AlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThr 351
 DB 1363 GCCACCCACCTTCGACGTCAGCGCGCCACCTTCGTGACCGCCACCTCTACTGGGAC 1422
 QY 352 AspAlaProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIle 371
 DB 1423 -----ACGGGCTCGAGCGACATCGACCTTACCTC 1452
 QY 372 ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsn 391
 DB 1453 TACACCCCAAC-----GGGACAGGTTGACTACTCTCTACCGCGCTAC 1497
 QY 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyr 411
 DB 1498 TAC-----GGCTTCGAGAGGTGGTCTACTTACCAACCGCGCGGAACTGG 1545
 QY 412 ThrIleGluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
 DB 1546 ACGGTCAAGGTCTGTCAGCTAC-----AAGGCGCGGGAACCTACAGGTGCGACGTC 1596
 QY 432 Val 432
 DB 1597 GTC 1599

RESULT 7
 AAX05926
 ID AAX05926 standard; DNA; 1977 BP.

XX AAX05926;
 AC
 XX
 DT 06-MAY-1999 (first entry)
 XX
 DE WO9856926 Seq ID 11.
 XX
 KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
 XX
 OS Thermococcus celer.
 XX
 PN WO9856926-A1.
 XX
 PD 17-DEC-1998.
 XX
 PF 04-JUN-1998; 98WO-JP002465.
 XX
 PR 10-JUN-1997; 97JP-00151969.
 XX
 PA (TAKI) TAKARA SHUZO CO LTD.
 XX
 XX Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
 XX WPI: 1999-080907/07.
 DR P-PSDB; AAW94840.
 XX
 XX Recombinant hyperthermostable protease from *Pyrococcus furiosus* - and
 PT gene encoding it, for large scale production of the protease for
 PT industrial use.
 XX
 PS Disclosure; Page 53-54; 82pp; Japanese.
 XX
 CC The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
 CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially *Bacillus* strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis
 XX
 SQ Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1-228-25 Length: 1977
 Score: 443.00 Matches: 137
 Percent Similarity: 46.03% Conservative: 66
 Best Local Similarity: 31.07% Mismatches: 170
 Query Match: 19.59% Indels: 68
 DB: 2 Gaps: 16
 US-09-985-689A-5 (1-433) x AAX05926 (1-1977)
 QY 8 ValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnValValAla 27
 DB 433 ATAGGGGCGGATACCGCTGTGGAACTCCCTCGGCTACGACGAAACGGGTGGTGGTGGC 492
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArg 47
 DB 493 ATCGTCGATACGGGTATAGACCGGAAC-----CACCCGATCTCGAAG 534
 QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
 DB 535 GGCAGAGTCATAGGTGGTGGTACGACCGCGCTCAACGCGAGGTGCGACCCCTACGATGACAC 594
 QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
 DB 595 GGACACGGAAACCCACGTTGCGGGTATCGTTGCGGGAACCGGACGCGTTAACTCCCGATAC 654

QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 100
 DB 655 ATAGCGTCCGCGCGCGGCGGAAGCTCGTCGGGTCAAGGTTCTCGGTCCGACGCTCG 714
 QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
 DB 715 GGAAGCGTCTCCACCATCATCGGGGTGTGTAGCTGGTGTGTCAGAACAAAGACAAAGTAC 774
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
 DB 775 GGGATAAGGGTCTATCAACCTCTCCCTCGGTCTCCCTCCAGAGTCCGACGGAACCGACTCC 834
 QY 141 SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaGly 160
 DB 835 CTCAGTCAGGCGCTCAACACGCTGGAGCGCGGTATAGTAGTCTGCGTCCGCGCGCGC 894
 QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 DB 895 AACAGCGGCGCGAACACCTACCGCTCGGTCTACCGCGCGCGCGAGCAAGGTCTATAACC 954
 QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
 DB 955 GTCCGTGCA-----GTTGACAGCAACGACAC 981
 QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 DB 982 ATCCCGAGCTTCTCCAGCAGGGACCGACCGCGAGGAGGCTCAAGCCGGAAGTCGTC 1041
 QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
 DB 1042 GCGCGCGGTGTGACATCATAGCCCGCGCGCGCAG-----GGAACGAGCATGGCG 1092
 QY 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
 DB 1093 ACCCGATAACAGCTACTACACCAAGCGCTCTGGAACCGAGCATGCCACCGCGACGTT 1152
 QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro--- 279
 DB 1153 TCGGCGGTGCGCGCTCATCTCCAG-----GCCACCGCGAGTGGACCGCGGAC 1203
 QY 280 -----LysProSerIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly--- 296
 DB 1204 AAGTGAAGACCGCGCTCATCGAGACCGCGAGCATAGTCGCCCGCCCAAGGATAGCGGAC 1263
 QY 297 LeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsn 316
 DB 1264 ATCGCTAC-----GGTCCGGTAGGTGAACGTCTACAAGGCC-----Lys 331
 QY 317 ValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGln-----Lys 331
 DB 1303 ATCAAGTACGACGACTACGCAAGCTCCTTCACCGGTCCGTCGCCACAGGGAAGC 1362
 QY 332 AlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThr 351
 DB 1363 GCCACCCACACTTCGACGTCAGCGCGCGCCACCTTCGTGACCGCCACCTCTACTGGAC 1422
 QY 352 AspAlaProGlySerThrAlaSerTyrThrLeuValAsnAspLeuValIle 371
 DB 1423 -----ACGGGTCCGAGCGACATCGACTCTTACCTTC 1452
 QY 372 ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsn 391
 DB 1453 TACGACCCCAAC-----GGGAACGAGTTGACTTCTCTACACCGCGCTAC 1497
 QY 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyr 411
 DB 1498 TAC-----GGCTTCGAGAGGTGGTCTACTACACCGCGCGCGGACCTGG 1545
 QY 412 ThrIleGluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
 DB 1546 ACGGTCAAGTTCGTGACGTAC-----AAGGGCGCGCGGAACTACCAAGTTCGACGTC 1596

```

QY 432 Val 432
DB 1597 GTC 1599

RESULT 8
AA05920
ID AAX05920 standard; DNA; 1236 BP.
XX AC AAX05920;
XX DT 06-MAY-1999 (first entry)
XX DE Hyperthermostable protease fragment encoding DNA.
XX KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
XX KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
XX OS Pyrococcus furiosus.
XX PN W09856926-A1.
XX PD 17-DEC-1998.
XX PF 04-JUN-1998; 98WO-JP002465.
XX PR 10-JUN-1997; 97JP-00151969.
XX PA (TAXI ) TAKARA SHUZO CO LTD.
XX PI Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
XX DR WPI; 1999-080907/07.
XX DR P-PSDB; AAW94836.
XX PT Recombinant hyperthermostable protease from Pyrococcus furiosus - and
XX PT gene encoding it, for large scale production of the protease for
XX PT industrial use.
XX PS Claim 6; Page 37-38; 82pp; Japanese.
XX CC The invention relates to a hyperthermostable protease derived from a
XX CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
XX CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
XX CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
XX CC 95 deg.C. The invention also provides gene sequences encoding a
XX CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
XX CC peptide from subtilisin, and PRO is the above protease. Host cells
XX CC (especially Bacillus strains) transformed with vectors comprising the
XX CC genes are used for the recombinant production of the protease. The
XX CC hyperthermostable protease which can be prepared in quantity suitable for
XX CC industrial use, can be used as an additive for drugs, washing agents and
XX CC foodstuffs and for chemical synthesis
SQ Sequence 1236 BP; 368 A; 251 C; 309 G; 308 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,89e-24 Length: 1236
Score: 422.50 Matches: 138
Percent Similarity: 43.96% Conservative: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.69% Indels: 79
DB: 2 Gaps: 17

US-09-985-689a-5 (1-433) x AAX05920 (1-1236)

QY 16 AsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
DB 58 AACTGGGATATGATGGTCTCGAATCAATAGGAATAATTCACACTGGAATTGAC--- 114
QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyGlySileThrAlaLeuTyrAla 55
DB 115 -----GCTTCTCATCCAGATCTCCAGGAAAGTA----- 144

```

```

56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
145 ATTGGGTGGGTAGATTGTCATGGTAGAGTATTCATACGATGACCATGACATGGA 204
70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
205 ACTCATGTAGCTTCAATAGCAGCTGCTACTGGAGCAGCAAGTAATGGCAAGTACAAGGA 264
85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
265 ATGGCTCCAGAGCTAAGCTGGCGGAAATTAAGGTTCTAGGTCCCGATGGTCTCGAAGC 324
103 LeuGlyGlyLeuProSerAsnLeuThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
325 ATATCTACTATATAATTAAGGAGTGTAGTGGCGGTTGATACCAAGATAAGTACGGAAT 384
123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
385 AAGGTCATTAAATCTTCTCTGTTCTCAAGCCAGAGCTCAGATGGTATCTCAGCGTCTAAGT 444
143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
445 CAGGCTGTTAATGCAGCGTGGGATGCTGGATTAGTTGTTGTTGTTGCGCTGGAAACAGT 504
163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
505 GGACCTAACCAAGTATACATCGGTTCTCCAGCAGCTGCAAGCAAGTATTATACAGTTGGA 564
183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla 202
565 CCGTTCGCAAGTAT-----GATGTTATAACA 591
203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProaspValThrAlaPro 222
592 AGCTTCTCAAGCAGAGGGCCAACTGCAGACGGCAGGCTTAAGCCTGAGGTTGTGTCTCA 651
223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyrAlaAsn 242
652 GGAACCTGGATAATTGCTCCAGAGCAAGT-----GGAACCTAGCATGGGTCAACCA 702
243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
703 ATTAATGACTATTACACAGCAGCTCTCGGACATCAATGGCAACTCTCTCAGCTAGTGGT 762
263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
763 ATTCCAGCCCTCTTGTCTCCAA-----GCACCCCGAGCTGGAGCTCCAGACAAAGTA 813
280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
814 AAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAAATAGCCGATATAGCC 873
299 TyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAla 318
874 TAC-----GGTGCAGGTAGGTTAATGCATACAGGCTATATAAC----- 912
319 TyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAla-----Thr 333
913 TACGATAACTATGCAAGACTAGTGTCTCCTGGATATCTGCCAACAAGGAGCAGCAACT 972
334 TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValThrThrAspAla 353
973 CACCAGTTCGTTATTAGCGAGCTTCGTTGCTACTCCACATTATATATATATATATATAT 1032
354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
1033 AAT-----AGCAGCTTGTATCTTACCTCTACGAT 1062
374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
1063 CCCAATGAAACACAGGTT-----GACTACTCTTAC-----ACC 1095
394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413

```

Db 1096 GCTACTATGGATTGCAAGAGTTGGTTATTACACCCCACTGATGGACATGGACAATT 1155
 QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
 Db 1156 AAGGTTGTAGTAC-----ACGGAAGTGCACAACTATCAAGTAGATGTGGTA 1203
 RESULT 9
 AAT85668
 ID AAT85668 standard; DNA; 1566 BP.
 XX AAT85668;
 AC
 XX
 DT 17-OCT-2003 (revised)
 DT 20-APR-1998 (first entry)
 XX
 DE Pyrococcus furiosus protease coding sequence.
 XX
 KW Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
 OS Pyrococcus furiosus; DSM-3638.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1566
 FT /tag= a
 FT /transl_except= (pos: 1282..1284, aa: Xaa)
 FT /note= "Xaa= Gly, Val"
 XX
 WO9721823-A1.
 XX
 PN 19-JUN-1997.
 PD
 XX
 PF 07-NOV-1996; 96WO-JP003253.
 XX
 PR 12-DEC-1995; 95JP-00323285.
 XX
 XX (TAKI) TAKARA SHUZO CO LTD.
 XX
 PI Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
 PI Tsunashawa S, Kato I;
 DR WPI; 1997-332794/30.
 DR P-PSDB; AAW24122.
 XX
 XX
 PT Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT industrially and as research reagents.
 XX
 PS Claim 7; Page 90-91; 159pp; Japanese.
 XX
 CC This sequence represents the coding sequence for the protease from
 CC Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents and
 CC industrially in the food, drug and chemical industries. (Updated on 17-
 CC OCT-2003 to standardise OS field)
 XX
 SQ Sequence 1566 BP; 467 A; 342 C; 372 G; 384 T; 0 U; 1 Other;
 Alignment Scores:
 Pred. No.: 3 86e-24 Length: 1566
 Score: 428.50 Matches: 138
 Percent Similarity: 43.96% Conservative: 55
 Best Local Similarity: 31.44% Mismatches: 167
 Query Match: 18.69% Indels: 79
 DB: 2 Gaps: 17
 US-09-985-689A-5 (1-433) x AAT85668 (1-1566)
 QY 16 AsnTyrGlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
 Db 58 AACTTGGGATGATGTTCTTGGAAATCAATAGGAATAATTGACACTGGATTGAC--- 114

QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
 Db 115 -----GCTTCTCATCCAGATCTCCAGGAAGATA----- 144
 QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
 Db 145 ATTGGGTGGGTAGATTGTGCAATGGTAGAGTTATCCATACATGACCATGGACATGGA 204
 QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
 Db 205 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGCGAAGTACAAGGA 264
 QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
 Db 265 ATGGCTCCAGAGCTAAGCTGGCGGAATTAAGGTTCTAGGTGCCGATGGTCTTGAAGC 324
 QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
 Db 325 ATATCTACTATTAATTAAGGAGTTGAGTGGCGCTTGATACCAAGATAAGTACGGAAT 384
 QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
 Db 385 AAGGTCAATTAATCTCTCTCTGTTCAAGCCAGAGCTCAGATGCTACTGACGCTCTAAGT 444
 QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
 Db 445 CAGGCTGTTAATGACGCTGGGATGCTGGATTAGTTGTTGTTGCTGCCGTGGAACAGT 504
 QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 Db 505 GGACCTAACCAAGTATACATCGGTTCTCCAGCAGCTGCAAGCAAGAGTTATTACAGTTGA 564
 QY 183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla 202
 Db 565 GCGCTTGACCAAGTAT-----GATGTTATAACA 591
 QY 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
 Db 592 AGCTTCTCAAGCAGAGGCCCACTGCAGACGCGAGGCTTAAGCCTGAGGTTGTTGCTCCA 651
 QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
 Db 652 GGAACCTGGATAATTGCTGCCAGAGCAAGT-----GGAACTAGCATGGGTCAACCA 702
 QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 703 ATTAATGACTATTACAGCAGCTCTCTGGACATCAATGGCAACTCTCTCAGTAGCTGT 762
 QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
 Db 763 ATTGAGCGCTCTTGTCTCCAA-----GCACCCCGAGCTGGACTCCAGCAAAAGTA 813
 QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
 Db 814 AAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAATATGCGATATAGCC 873
 QY 299 TyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 318
 Db 874 TAC-----GGTGCAGTAGGTTAATGCATACAGGCTATAAAC----- 912
 QY 319 TyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAla-----Thr 333
 Db 913 TACGATAACTATGCAAGCTAGTTGTTCACTGGATATGTTGCCAACAAAGGCGAGCAAACT 972
 QY 334 TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAla 353
 Db 973 CACCAGTTCTGTTATTAGCGGAGCTCGTTCGTAATCCACATTATATCTGGGACAAATGCC 1032
 QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
 Db 1033 AAT-----AGCGACCTTGATCTTTACCTCTACGAT 1062
 QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393

394	QY	Gly	Arg	Asn	Asn	Val	Glu	Asn	Val	Phel	Asn	Ala	Pro	Gln	Ser	Gly	Thr	Thr	Leu	413
1492	DB	GCT	TAC	TAT	CGA	ATC	GAA	AGT	TGG	TAT	TAC	ACC	CAC	TGA	TGA	AA	CA	TGC	AAAT	1551
414	QY	Glu	Val	Gln	Ala	Tyr	Asn	Val	Pro	Ser	Gly	Pro	Gln	Arg	Phe	Ser	Leu	Ala	Leu	432
1552	DB	AAG	GTG	TAA	GC	TAC	-----	AGC	GAA	GTC	CAA	CA	ACT	CA	AG	TAG	ATG	TG	TGA	1599

RESULT 11

AAX05929

ID AAX05929 standard; DNA; 1962 BP.

XX

AC AAXC

XX

DT 06-1

XX

DE "HYPE"

XX
LVIKW
HYPER-
=dd-

XY
MY

50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000

XX 33 Fyfe

PN
W091

XX

PD 17-1

XX.

PF 04-1

XX

PR 10-

XX 23

PA (TA)

XX	DT	TX
----	----	----

PI YY

AA DR WPT

DR	WET
DR	P-P

XX
XX

Rec

PT gen

ind

XX

PS Dis

XXIX

CC The

the

22
23
24

2007

2003

22
23

200

CC
gen

22 23

cc
ind

QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
 Db 1459 CCCAATGGAAACCAAGTT-----GACTACTCTTAC-----ACC 1491
 QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413
 Db 1492 GCCTACTATGATTCCAAAGAGTTGGTTATTACAAACCACTGATGGAACTGACCAATT 1551
 QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
 Db 1552 AAGTTGTAAGCTAC-----AGCGGAAGTGCAAACTATCAAGTAGATGTGGTA 1599

RESULT 12

AAT85669
 ID AAT85669 standard; DNA; 1977 BP.
 XX
 AC AAT85669;
 DT 20-APR-1998 (first entry)
 XX
 DE Protease coding sequence.
 XX
 KW Protease; research reagent; thermal stability; ss.
 XX
 OS Synthetic.
 XX
 PN WO9721823-A1.
 XX
 PD 19-JUN-1997.
 XX
 PF 07-NOV-1996; 96WO-JP003253.
 XX
 PR 12-DEC-1995; 95JP-00323285.
 XX
 PA (TAKI) TAKARA SHUZO CO LTD.
 XX

PI Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
 PI Tsunawasa S, Kato I;
 XX
 DR WPI; 1997-332794/30.
 XX
 DR P-PSDB; AAW24123.

XX
 PT Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT industrially and as research reagents.

XX
 PS Claim 11; Page 95-97; 159pp; Japanese.

XX This sequence represents the coding sequence for a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries

XX
 SQ Sequence 1977 BP; 527 A; 562 C; 513 G; 375 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,36e-22 Length: 1977
 Score: 401.50 Matches: 130
 Percent Similarity: 43.96% Conservative: 63
 Best Local Similarity: 29.61% Mismatches: 167
 Query Match: 17.76% Indels: 79
 DB: 2 Gaps: 17

US-09-985-689A-5 (1-433) x AAT85669 (1-1977)

QY 16 AsnTyrGlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
 Db 454 AACTTGGGATATGATGGTCTTGGAAATCAATAGGAATAATTGACACTGGAATTGAC--- 510
 QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
 Db 511 -----GCTTCTCATCCAGATCTCCAGGAAAGTA----- 540

QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
 Db 541 ATTGGTGGGTAGATTCTCAATGGTAGAGTTATCCATAGATACCATGACATGGA 600
 QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
 Db 601 ACTCATGTAGTTCATATAGCAGCTGTACTGGAGCAGCAAGTAATGGCAAGTACAAGGA 660
 QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
 Db 661 ATGGCTCCAGAGCTAAGCTGGCGGAATTAAGTTCTAGGTGCCGATGGTCTCTGGAAGC 720
 QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
 Db 721 ATATCTACTATATTAAGGAGTTGAGTGGCGGTGATACAAAGATAAGTACGGAATT 780
 QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
 Db 781 AAGGTCAATTAATCTTTCTCTGTGTTCAAGCCAGAGCTCCGACGGAACCGACTCCCTCA 840
 QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
 Db 841 CAGCGCTCAACACCGCTGGAGCGGTATAGTAGTCTGCGTCGCCCGCGGCAACAGC 900
 QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 Db 901 GGGCCGAACACCTACACCTCGCTCGCTCACCGCGCGAGCAAGGTCTATAACCGTCGT 960
 QY 183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla 202
 Db 961 GCA-----GTTGACAGCAACGACACATCGCC 987
 QY 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
 Db 988 AGCTTCTCCAGAGCGGACCGACCGCGGACGGAAGGTCTAAGCGGAAGTCTGTCGCCCC 1047
 QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPheTrpAlaAsn 242
 Db 1048 GCGCTTGACATCATAGCCCCCGCGCGCAGC-----GGACACGATGGCACCCCG 1098
 QY 243 TyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProfileValAlaGly 262
 Db 1099 ATAACGACTACTACACCAAGGCTCTGGAACCCAGCATGCCACCCGCGACGTTTCGGGC 1158
 QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
 Db 1159 GTTGGCGCGCTCATCTCTCCAG-----GCCACCCGAGTGGACCCCGGCAAGGTG 1209
 QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
 Db 1210 AAGACCCCTCATCGAGACCGCGCATAGTCGCCCCCAAGGATAGCGGACATCGCC 1269
 QY 299 TyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 318
 Db 1270 TAC-----GGTGGCGGTAGGTGAACGTCTACAAGGCC-----ATCAAG 1308
 QY 319 TyrValAsnGluAlaThrAlaLeuAlaThrGlyGln-----LysAlaThr 333
 Db 1309 TAGACGACTACGCCAAGCTCACTTCACCGGTCCGTCGCCACCAAGGAAGCGCAC 1368
 QY 334 TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAla 353
 Db 1369 CACACCTTCGAGTCAGCGCGCCACCTTCGTACCGGCCACCTCTACTGGGAC----- 1422
 QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
 Db 1423 -----ACGGGCTCGAGGACATCGACTCTTACTCTCTAGGAC 1458
 QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
 Db 1459 CCCAAC-----GGGAACGAGTTGACTACTCTCTACACCGGCTACTAC--- 1500
 QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413

```

Db      1501  -----GGCTTCGAGAGGTCGGCTACTCAACCCGACCGCGGAACCTGGACGTC 1551
Qy      414  GluValGlnAlaTy rAsnValProSerGlyProGlnArgPheSerLeuAla rVal 432
Db      1552  AAGGTCGTCAGCTAC-----AAGGGCGCGGAACCTACGAGTCGACGTC 1599

RESULT 13
ABLS4900
ID      ABL54900 standard; DNA; 2121 BP.
XX
XX      AC      ABL54900;
XX
XX      11-SEP-2003 (revised)
XX      31-MAY-2002 (first entry)
XX
XX      T. yonsei subtilisin-like serine protease coding sequence.
XX
XX      Subtilisin-like serine protease; ss.
XX
XX      Thermoanaerobacter yonseiensis.
XX
XX      Key      Location/Qualifiers
XX      CDS      142..1779
XX              /*tag= a
XX              /product= "subtilisin-like serine protease"
XX
XX      KR2000072:41-A.
XX
XX      05-DEC-2000.
XX
XX      04-AUG-2000; 2000KR-00045411.
XX
XX      04-AUG-2000; 2000KR-00045411.
XX
XX      (KIMY/) KIM Y S.
XX
XX      Chang HJ, Kim DH, Byun YR, Kim YS;
XX      WPI: 2001-298092/31.
XX      P-PSDB; AB509483.
XX
XX      New DNA sequence of thermophilic protein decomposition enzyme and protein
XX      derived therefrom.
XX
XX      Claim 1; Page 6; 15pp; Korean.
XX
XX      This sequence represents the DNA encoding the Thermoanaerobacter yonsei
XX      subtilisin-like serine protease of the invention. (Updated on 11-SEP-2003
XX      to standardise OS field)
XX
XX      Sequence 2121 BP; 712 A; 410 C; 425 G; 554 T; 0 U; 20 Other;

Alignment Scores:
Pred. No.:      4,246-20      Length:      2121
Score:          373.50      Matches:      132
Percent Similarity: 45.85%      Conservative: 56
Best Local Similarity: 32.20%      Mismatches: 139
Query Match:    16.52%      Indels:      83
DB:             4      Gaps:      18

US-09-985-689A-5 (1-433) x ABL54900 (1-2121)

Qy      6  GlyTieValIy eAlaAspValAlaGlnAsnAsnTy rGlyLeuTy rGlyGlnGlyGlnVal 25
Db      445  GGAATCACAAA-----GCAGGAGTCATTTGGAGTCACAGGAAAAAATATAACA 495
Qy      26  ValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerVethisGluAla 45
Db      496  ATAGCAATATTTCACACAGGTATACACGGAATCACGTTGACCTCTCA----- 543
Qy      46  PheArgGlyIy sileThrAlaLeuTy rAlaLeuGly rArgThrAsnAsnAlaAsn----- 63

```


QY 385 SerTyrProTyrAspAsnAsnTrpAspGlyArg-----AsnAsnValGlu 399
 Db 2048 AGCGCTCGTGCATGATGACCGCGGACACCGGCTCGGGGCGACGGTGCACGCGCGTAC 2107
 QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle----- 413
 Db 2108 TCGCGGTAGTGTGTCGCC---ACGGGCGGCGGCGACAGCGTCCGACGCGCCCGCGGTG 2164
 QY 414 -----GluValGlnAlaTyrAsnVal-ProSerGly 423
 Db 2165 CAGCGGAGTGCAGTCTGACGACGTCGACGTCGTCGCGC 2201

RESULT 15

AAAT61455
 ID AAT61455 standard; DNA; 2809 BP.

XX AC AAT61455;

DT 17-OCT-2003 (revised)

DT 06-OCT-1997 (first entry)

XX DhpA-mel chimeric gene.

XX asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;
 KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
 KW cardiovascular; treatment; hypertension; ischaemic heart disease; ds.

XX Streptomyces viridosporus.

OS Streptomyces antibioticus.

OS Chimeric.

PH Key Location/Qualifiers

FT CDS 338..2809

FT /*tag= a

FT mat_protein 338..2539

FT /*tag= b

FT /product= "DhpA protein product"

FT /note= "from S. viridosporus"

FT mat_protein 2540..2809

FT /*tag= C

FT /product= "Melanin"

FT /note= "from S. antibioticus"

XX WO9705243-A1.

XX 13-FEB-1997.

XX 30-JUL-1996; 96WO-JP002147.

XX 31-JUL-1995; 95JP-00212975.

XX 29-FEB-1996; 96JP-00067478.

XX (SAOC) MERCIAN CORP.

XX Arisawa A, Matsufuji M, Tsuruta T, Dobashi K, Nakashima T;

XX Ishihiki K, Yoshiohka T;

XX WPI; 1997-145682/13.

XX P-PSDB; AAW13668.

XX Asymmetric hydrolase gene derived from Streptomyces viridosporus - acts

PT on 4-substituted-1,4-di:hydro:pyridine derivatives to produce chiral

PT derivatives useful for synthesis of cardiovascular drugs.

XX Claim 5; Page 37-43; 78pp; Japanese.

XX This sequence is a fusion gene encoding Streptomyces viridosporus dhpA

CC gene product (an asymmetric hydrolase) which acts on 4-substituted-1,4-

CC dihydropyridine derivatives, and melanin from S. antibioticus. The dhpA

CC enzyme allows the efficient conversion of 4-substituted-1,4-

CC dihydropyridine esters to chiral partially hydrolysed derivatives, for

CC use in the synthesis of cardiovascular drugs suitable for the treatment

CC of e.g. hypertension and ischaemic heart disease. (Updated on 17-OCT-2003
 CC to standardise OS field)
 XX Sequence 2809 BP; 473 A; 1041 C; 984 G; 311 T; 0 U; 0 Other;

Alignment Scores:

Pred. NO.: 9.83e-18 Length: 2809
 Score: 345.50 Matches: 133
 Percent Similarity: 41.06% Conservative: 53
 Best Local Similarity: 29.36% Mismatches: 169
 Query Match: 15.28% Indels: 98
 DB: 2 Gaps: 19

US-09-985-689A-5 (1-433) x AAT61455 (1-2809)

QY 18 GlyLeuTyrGlyGlnGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
 Db 998 GGCTACGACGCAAGGCGGTGAAGATCGCGTCTCTGGACACCGCGTGTGACACGAGC--- 1054
 QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyValIleThrAlaLeuTyrAlaLeuGly 57
 Db 1055 -----CATCCGGACCTGAGGGCGGGGTGACCGCGTCCAGAACTTCACC 1099
 QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
 Db 1100 GCCCGCGCGCGCGCGCGGACAGGTGGCGCACCGGACCCACCGTCTCGATCGCGGCG 1159
 QY 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
 Db 1160 GGCACGGCGGCCAGTCCAGGGCAAGTACAAGGGCGTGCACCCCGCGCGCGGATCTC 1219
 QY 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThr 112
 Db 1220 AACGGCAAGGTCTCTCGACGACTCC-----GGTTTCGGCGACGACTCCGGCATC 1267
 QY 113 LeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHisThrAsnSerTrpGly 130
 Db 1268 CTCGCGCGCATGTGAGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1327
 QY 131 Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
 Db 1328 GGCATGGACACACCGGAGACCGCGCTGTGAGGCGCG----- 1366
 QY 144 ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly 163
 Db 1367 GTCGACAAGCTGTCCGCGCGAGAGGGGTCTCTGTTCCTCCATTCGCGCGCGCGCAAGGG 1426
 QY 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
 Db 1427 CCGGAGTCC-----ATCGGTTCGCGGACGCGCGCGCGCGCGCGCGCGCGCGCG 1480
 QY 184 ThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAlaGln 203
 Db 1481 -----GTCGACGACAAAGGCAAGCTCGCGCAC 1507
 QY 204 PheSerSerArgGly-----AlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
 Db 1508 TTCTCTCTCACCGGCCCGCGCTCGGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1567
 QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsn 242
 Db 1568 GCGGTGGACATCACGCGCGCTCGCGGAGGGGCAACGACATCGCGCGAGGAGGTGGTGAG 1627
 QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 1628 GGACCGCGCGGTATCATGACCATCTCCGCGACGTGATGGGCGACCGCGCGCGCGCGCG 1687
 QY 263 AsnValAlaGlnLeuArgGluPheIleLysAsnArgGlyIleThrProLysProSer 282
 Db 1688 GCGCGCGCGCTCTGACGACGAG-----CACCCCGAC 1720
 QY 283 LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr 299

```
Db 1721 TGGACCTCGCGCACTGAAGGGCGCGCTCACC GGCTCCACCAAGGGCGGC---AAGTAC 1777
Qy 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1778 ACCCGGTTCGAGCAGGCTCGGGCGGATCCAGCCCGACAAGCGCTCCAGCAGACCGTG 1837
Qy 320 ValAsnGluAlaThrAlaLeuAlaThrGly----- 329
Db 1838 ATCCCGCACC CGGTCTCGGTAGCTTCGGGTCCAGCAGTGGCGGCACACCGACGAGGAG 1897
Qy 330 -----GlnLysAlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIle 346
Db 1898 CCGGTCAACCAAGCAGCTGACCTACCGCAACCTCGGCACCCAGGACGTCACGCTGAAGCTG 1957
Qy 347 SerLeuValTyrThrAspAlaProGlySerThrThrAlaSer-----TyrThrLeuVal 364
Db 1958 ACGTCGACCGCACCGACCCCAAGGCAAGGCGGCCCGGGGGCTTTCACGCTGGGC 2017
Qy 365 AsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPhe 384
Db 2018 -----GCCACCAACCGTGACCGTCCCGCGGGCGGC----- 2047
Qy 385 SerTyrProTyrAspAsnAsnTyrAspGlyArg-----AsnAsnValGlu 399
Db 2048 AGCGCTCCGTCGACATACCGCCGACACCCGGTTCGGCGGCACGGTGGAGCGCGGTAC 2107
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle----- 413
Db 2108 TCGGGTACGTGTGCGCC---ACGGGCGGGCGGCAGACGGTCCGCAAGCGCGCGCGGTG 2164
Qy 414 -----GluValGlnAlaTyrAsnVal-ProSerGly 423
Db 2165 CAGCGCAGGTGAGTCTGACGACGTACCGTCCGGC 2201
```

Search completed: April 4, 2004, 01:12:23
Job time : 362.908 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:50:44 ; Search time 63.4512 Seconds
(without alignments)
3787.066 Million cell updates/sec

Title: US-09-985-689A-5
Perfect score: 2261
Sequence: 1 NDVARGIVKADVAQNNVGLY.....EVQAYNVPSQPFSLAIVH 433

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09985689/runat 31032004 161807 4184/app query.fasta_1.3498
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THE MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 @CGN 1.1.142 @runat 31032004 161807 4184 -NCFU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DISPLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2044	90.4	3003	2	US-08-873-479-41
2	2002.5	88.6	1920	4	US-09-509-814A-3
3	1999.5	88.4	1923	4	US-09-509-814A-7
4	1998.5	88.4	1923	4	US-09-509-814A-5
5	443	19.6	1977	3	US-08-894-818B-2
6	443	19.6	1977	4	US-09-445-472-11
7	422.5	18.7	1236	4	US-09-445-472-2
8	422.5	18.7	1566	3	US-08-894-818B-4
9	422.5	18.7	1962	3	US-08-894-818B-34
10	422.5	18.7	1962	4	US-09-445-472-15
11	401.5	17.8	1977	3	US-08-894-818B-6
12	350.5	15.5	2539	3	US-09-000-016-3

13	350.5	15.5	2539	4	US-09-514-340-3	Sequence 3, Appli
14	345.5	15.3	2809	3	US-09-000-016-1	Sequence 1, Appli
15	345.5	15.3	2809	4	US-09-514-340-1	Sequence 1, Appli
16	312	13.8	2632	1	US-07-671-376C-4	Sequence 4, Appli
17	307	13.6	2835	1	US-08-750-532-2	Sequence 2, Appli
18	307	13.6	4765	1	US-08-750-532-6	Sequence 6, Appli
19	307	13.6	4765	3	US-08-894-818B-7	Sequence 7, Appli
20	307	13.6	4765	4	US-09-445-472-5	Sequence 5, Appli
21	306.5	13.6	1859	3	US-08-894-818B-15	Sequence 15, Appli
22	283	12.5	898	1	US-08-750-532-7	Sequence 7, Appli
23	282	12.5	564	1	US-08-750-532-11	Sequence 11, Appli
24	282	12.5	564	3	US-08-894-818B-14	Sequence 14, Appli
25	269	11.9	1140	1	US-08-322-677A-11	Sequence 11, Appli
26	269	11.9	1140	1	US-08-322-676-11	Sequence 11, Appli
27	269	11.9	1140	3	US-08-898-218-11	Sequence 11, Appli
28	269	11.9	1140	3	US-08-848-793-11	Sequence 11, Appli
29	269	11.9	1140	4	US-09-445-270-6	Sequence 6, Appli
30	269	11.9	1140	4	US-08-322-678-11	Sequence 11, Appli
31	269	11.9	1143	3	US-08-269-050-1	Sequence 1, Appli
32	269	11.9	1143	6	5336611-1	Patent No. 5336611
33	269	11.9	1320	6	5217878-1	Patent No. 5217878
34	269	11.9	1350	3	US-09-049-867-1	Sequence 1, Appli
35	269	11.9	2280	1	US-07-661-378A-1	Sequence 1, Appli
36	269	11.9	2280	1	US-07-918-318-1	Sequence 1, Appli
37	269	11.9	2280	2	US-08-413-724-1	Sequence 1, Appli
38	269	11.9	2280	3	US-08-853-494-1	Sequence 1, Appli
39	269	11.9	10216	2	US-08-875-154-1	Sequence 1, Appli
40	268	11.9	807	1	US-07-706-691G-72	Sequence 72, Appli
41	268	11.9	807	2	US-08-254-021-72	Sequence 72, Appli
42	268	11.9	807	2	US-08-618-446-72	Sequence 72, Appli
43	268	11.9	807	3	US-08-980-135-72	Sequence 72, Appli
44	268	11.9	807	4	US-09-585-798-72	Sequence 72, Appli
45	268	11.9	1140	1	US-08-322-677A-12	Sequence 12, Appli

ALIGNMENTS

RESULT 1

Sequence 41, Application US/08873479

Patent No. 5891701

GENERAL INFORMATION:

APPLICANT: Sloma, Alan

APPLICANT: Lynne, Christianson

TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide

TITLE OF INVENTION: Having Protease Activity

NUMBER OF SEQUENCES: 57

CORRESPONDENCE ADDRESS:

ADDRESSEE: NO. 5891701o No. 5891701disk of No. 5891701th America

STREET: 405 Lexington Avenue

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10174

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/873,479

FILING DATE: 12-JUN-1997

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Agria, Cheryl H

REGISTRATION NUMBER: 34,086

REFERENCE/DOCKET NUMBER: 5251.000-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

TELEX:

INFORMATION FOR SEQ ID NO: 41:

SEQUENCE CHARACTERISTICS:

US-09-985-689A-5 (1-433) x US-09-509-814A-3 (1-1920)

QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 616 AATGATGTGGCAGAGGATTTGTCAAGCGGATGTGGCAGCAGCAGCTACGGTTTGTAT 675
QY 21 GlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyValAsnAspSer 40
Db 676 GGCAAGGCCAGATTGTGCGAGTTCGCCGATCTGGATTGGATACAGGAAGAAACGACAGT 735
QY 41 SerMetHisGluAlaPheArgGlyIysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGTCATGAAGCCCTTCGGCGTAAAAATAACAGCACTATATGCACTGGGTGGCAGCGAAT 795
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 796 AATGCGATGATACGACCGTCATGGTACCCATGTGGCAGGTTCGGTATTAGGAATGGC 855
QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 856 GCAACGAATAAAGGAATGCACTCAAGCGAATCTGGTTTTTCAATCCATCATGATGATGC 915
QY 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 916 AGTGTGGGCTTGGAGGCTTGGCTTCCAAATCTGCAAACTTATTCAGGCCAAGCAATTCAGT 975
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValAlaThrAla 139
Db 976 GCAGGTGCCAGAAATTCATAACAACCTCTGGGGGGCAGCGGTGAATGGGGCTCACAGACA 1035
QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValIlePheAlaAla 159
Db 1036 GATTCAGAAATGTGGATGACTATGTAGGAAATGATATGACGATCTTTTTCGGGCT 1095
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1096 GGGATGAAAGCCGACCGCGTACCATCACTGCACTGTCAGCTGGTACGGCTAAAAACGCCATA 1155
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
Db 1156 ACAGTGGCGCAACCGAAACCTGGCTCCAGCTTCGGTTCCTATGCAGATAATATTAAAC 1215
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyValGlyIleLeuProAspVal 219
Db 1216 CAGGTGACAGCTTCTTCCGTGGCCGACAAAAGATGGCGGATCAAGCCTGATGTC 1275
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 1276 ATGGCCCGCAGGACATACATTTTATCAGCAGATCTTCTTTCACCCGATTCCTCCTTC 1335
QY 240 TrpAlaAsnTyrAsnSerIysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
Db 1336 TGGCGAATCATGACAGCAAAATATGCCTATATGGTGGNACGTCCTATGGCAACCCGATT 1395
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 1396 GTTGGGGGAATGTTGCACAGCTCCGTGAGCATTTTGTGAAAATATAGAGGAATCACTCCT 1455
QY 280 LysProSerLeuIleLysAlaAlaIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1456 AAGCCTTCCCTATATGAAGCAGCTTGAATGCAAGGTGCTGCTGATTTGGATTGGGTTAT 1515
QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1516 CCGAACCGAAACCAAGGATGGGCGGAGTGACCTGGATAAATCGTTGAACGCTGCCTAT 1575
QY 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
Db 1576 GTGAACGGAATCCCACTCCCTCACTAGCCAAAAGCGACATATACCTTTTACTCAACG 1635
QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAla 359
Db 1636 CGGGGCAAGCCATGAAATCTCCCTGGTATGGTCCGATGCCCTCGCAAGCACTACTGCT 1695

QY 360 SerTyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1696 TCTGTAAACCTGGTCAATGATTGGATTGGTTCATTACAGCACCAACGGAACAGATAT 1755
QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
Db 1756 GTCGGGAATGACTCTTCAGCACCATTTGCAATAACTGGGATGGCGCAATTAACGTAGAA 1815
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1816 AATGTAATTATTAATTCGCCCCAAAGTGGAAACATATACCATTCAGGTGCAAGCATATAT 1875
QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1876 GTCCCGTGGACACAAAACCTTCTGTTGGCAATTTGTGAAC 1917
RESULT 3
US-09-509-814A-7
; Sequence 7, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUM
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-OPCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-7
Alignment Scores:
Pred. No.: 2,34e-190 Length: 1923
Score: 1999.50 Matches: 381
Percent Similarity: 94.24% Conservative: 28
Best Local Similarity: 87.79% Mismatches: 24
Query Match: 88.43% Indels: 1
DB: 4 Gaps: 1
US-09-985-689A-5 (1-433) x US-09-509-814A-7 (1-1923)
QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 619 AATGATGTGGCGGTGGAAATTCAAAGCGGATGTGGCTCAGACAGCTACGGGTTGTAT 678
QY 21 GlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGCAAGCACAGATCGGTAGCGGTTCGGATACAGGGCTTGATACAGCTCGCAATGACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 739 TCGATGTCATGAAGCCCTTCCCGGGGAAATTAATCTGCAATTAATTCGATTCGACCGACGAAT 798
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 799 AATGCCAATGATACGAATGCTGATGGTACGCATGTGGCTCGCTCCGTATTAGGAAACGGC 858

RESULT 4
US-09-509-814A-5
; Sequence 5, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; PCT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 5
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-5
Alignment Scores:
Pred. No.: 2,94e-190 Length: 1923
Score: 1998.50 Matches: 381
Percent Similarity: 94.24% Conservative: 28
Best Local Similarity: 87.79% Mismatches: 24
Query Match: 88.39% Indels: 1
DB: 4 Gaps: 1
US-09-985-689A-5 (1-433) x US-09-509-814A-5 (1-1923)
QY 1 AsnAspValAlaAtrGlyLeuValValAlaAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 619 AATGATGTTGGCGTGAATTTGTCAGAGCGGATGTGGCTCAGAGCAGCTCGGGTTGAT 678
QY 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyValAsnAspSer 40
Db 679 GGACAAGGACAGATCGTAGCGGTTCCCGATACAGGGCTTGATACAGGTTCGCAATGACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 739 TCGATGTCATGAAGCCTTCGGCGGGAATTTACTGCAATTATATGCAATTCATCGGACGGAAT 798
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaAlaGlySerValLeuGlyAsn--- 79
Db 799 AATGCCAATGATACCGAATGGTTCATGTCAGCATGTGGCTGGCTCGCTTCGTTATTAGGAACGCGC 858
QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 859 TCCACTAATAAAGGAATGCGGCTCAGGCGAATCTAGTCTTCCAATCTATCATGTGGATAGC 918
QY 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 919 GTGGGGGACTTGGAGGACTACTTCGGAATCTGAAACCTTATTTCAGCCAGCATACAGT 978
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAla 139
Db 979 GCTGGTGCAGAAATTCATACAACTCTCGGAGCAGCATGTAATGGGGCTTACACAACA 1038
QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 1039 GATTCCAGAAATGTGGATGACTGTGGCGCAAAATGATATGACCATCTTTTTCGCTGCC 1098

QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 859 TCCACTAATAAAGGAATGCGGCTCAGGCGAATCTAGTCTTCCAATCTATCATGTGGATAGC 918
QY 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 919 GTGGGGGACTTGGAGGACTACTTCGGAATCTGCAAACTTATTTCAGCCAGCATACAGT 978
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAla 139
Db 979 GCTGGTGCAGAAATTCATACAACTCTCGGAGCAGCATGTAATGGGGCTTACACAACA 1038
QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 1039 GATTCCAGAAATGTGGATGACTGTGGCGCAAAATGATATGACCATCTTTTTCGCTGCC 1098
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1099 GGCAATGAAGGACCGAAGCGCGAACCATCAGTGCACAGGACAGCTAAATAATGCAATA 1158
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
Db 1159 ACAGTGGAGCTACGGAACCTCCGCCCAAGCTTTGGGTCTTATCGGACATATCAAC 1218
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 1219 CATGTGCACAGTTCCTTCACTGGACCGACAAAGGATGGAGCGATCAAAACCGGATGC 1278
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 1279 ATGGCACCGGAGCTTCATACATATCAGCAAGATCTCTCTGCACCGGATTCCTCCCTC 1338
QY 240 TrpAlaAsnTyrAsnSerLysTrpAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
Db 1339 TGGCGGACCATGACAGTAATATGATACATGCGTGAAGCTCCATGGCTACACCGATC 1398
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 1399 GTTCTGGAACCGTGGACAGCTTCGTGAGCATTTTGTGAANAACAGAGCATCACACCA 1458
QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1459 AAGCCTTCTCTATTAAAGCGGCACTGATTCGCGGTGCAGCTGACATCGGCTTGGCTAC 1518
QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1519 CCGAACCGTAACCAAGGATGGGACGAGTGACATTTGGATATAATCCCTGAAACGTTGCCAT 1578
QY 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
Db 1579 GTGAACGAGTCCAGTCTCTATCCACGCAAAAGCGAGCTACTCGTTTACTGCTACT 1638
QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAla 359
Db 1639 GCCGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCCCTCGGAGCACAACCTGCT 1698
QY 360 SerTyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1698 TCCGTAACGCTTGTCAATGATCTGGACCTTGTCTATCCGCTCCCAATGGCACACAGTAT 1758
QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
Db 1759 GTAGGAATGACTTTTACCTCGCATACAAATGATAACTGGGATGGCGCGCAATAAGCTAGAA 1818
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1819 AATGTATTATTAAATGCACCAAAAGCGGAGCATATCAATTAAGTACAGGCTTATAC 1878
QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1879 GTACCGGTTGGACACAGAACTTCTCGTTGGCAATTGTGAAT 1920

```

QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1099 GGGAAATGAAGACCCAGCGGGAACCATCATGTCACCGAGCACAGCTAAATAATGCAATA 1158
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
Db 1159 ACAGTCGGAGCTACGGAAACCTCGCCAGCTTTGGTCTTATGCGGCAATATCAAC 1218
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 1219 CATGTGGCACAGTCTCTTCACTGGACCGCAACAGAGTGGACGGATCAAAACCGGATGTC 1278
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 1279 ATGGCACCGGAACGTTCACTATCATACAGAGATCTTCTGTGACCGGATCTCTCTTC 1338
QY 240 TrpAlaAsnTyrAsnSerIleTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
Db 1339 TGGCGGAACCATGACAGTAAATATGATCATATGGGTGGAACTCCATGGCTACACCGATC 1398
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 1399 GTTGCTGGAACCGTGGACAGCTTCGTGAGCATTTTGTGAAACAGAGGATCACACA 1458
QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1459 AAGCCTTCTATTAAAGCGGCACCTGATTGCGGTGCGAGTGACATCGGCTTGGCTAC 1518
QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1519 CCGAACGGTAACCAAGATGGGAGCGAGTGACATGGATTAATCCCTGAACGTTGCCAT 1578
QY 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
Db 1579 GTGAACGAGTCCAGTCTCTATCCACCGCAACCAAGCGAGTACTCGTTTACTGCTACT 1638
QY 340 AlaGlyLysProLeuLysIleSerLeuValThrThrAspAlaProGlySerThrThrAla 359
Db 1639 GCGCGAAGCCTTTGAAATCTCCCTGGTATGCTGTGATGCCCTCGGAGCAACTGCT 1698
QY 360 SerTyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1699 TCGGTAAAGCTTGCTAATGATCTGGACCTTGTCATTCGCTCCAAATGGCACAGTAT 1758
QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnThrAspGlyArgAsnAsnValGlu 399
Db 1759 GTAGGAAATGACATTACTTCGCCATACATGATACTGGGATGGCGGCAATAACGTAGAA 1818
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrIleGluValGlnAlaTyrAsn 419
Db 1819 AATGTATTTATTAATGACCAACAAAGCGGACGATATCAATTTGAGGTACAGGCTTATAAC 1878
QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1879 GTACCGGTTGGACACAGACCTTCTCGTTGGCAATGTGTAAT 1920

```

RESULT 5

```

US-08-894-818B-2
; Sequence 2, Application US/08894818B
; Patent No. 6261822

```

GENERAL INFORMATION:

```

; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshi
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark

```

```

; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-894-818B-2

```

Alignment Scores:

```

Pred No.: 1,4e-34 Length: 1977
Score: 443.00 Matches: 137
Percent Similarity: 46.03% Conservative: 66
Best Local Similarity: 31.07% Mismatches: 170
Query Match: 19.59% Indels: 68
DB: 3 Gaps: 16

```

```

US-09-985-689A-5 (1-433) x US-08-894-818B-2 (1-1977)

```

```

QY 8 ValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnGlyGlnValAla 27
Db 433 ATAGGGGCGGATACCGTCTCGAACTCCCTCGGTACCGACGAGCGGTGTGGTGGCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATCGTCGATACGGGTATAGACGGAAC-----CACCCCGATCTCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
Db 535 GGCAGGTCTATAGGCTGTGTACGACGCGGTCAACGGCAGGTGCGACCCCTCATGATGACCAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
Db 595 GGACACGGAACCCAGCTTGGGGTATCGTTGCCGGAACCGGACGCGGTAACTCCAGTAC 654
QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 100
Db 655 ATAGCGGTGCGGCGCGGGAAGCTGTCGCGCGTCAAGGTTCTCGGTGCGGCGGTTCG 714
QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
Db 715 GGAAGCGTCTCCACCATCATCGCGGTGTTCAGTGGGTCTGTCAGAGAACAGACAGTAC 774
QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
Db 775 GGGATAAGGTCATCAACCTCTCCCTCGGCTCTCCAGAGTCCCGACGGAACCGACTCC 834

```

141 SerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAagly 160
 835 CTCAGTCAGCCGCTCAACACCCCTGGAGCGCGGTATAGTAGTGGTGGTCCGCCGC 894
 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLeuAsnAlaIleThr 180
 895 AACAGCGGGCCGGAACACCTACAGCTCGCTCCAGCGCGCGCGGCAAGAGTGCATAACC 954
 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
 955 GTCGGTGCA-----GTTGACAGCAACGACAC 981
 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLeuProAspValThr 220
 982 ATCGCCAGCTTCTCCAGCAGGGGACCGACCGCGGACCGAAGCTCAAGCCGGAAGTCTGTC 1041
 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
 1042 GCCCGCGGCTTGACATCATAGCCCGCGCGCAGC-----GGACACGATGGGC 1092
 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
 1093 ACCCGATAAAGACTACTACACCAAGGCTCTGGAACCCAGCATGGCCACCCCGACGCTT 1152
 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro--- 279
 1153 TCGGGCGTGGCGGCTCATCTCCAG-----GCCACCCGAGTGGACCCCGGAC 1203
 280 -----LysProSerLeuIleLysAlaLeuAlaGlyAlaThrAspValGly--- 296
 1204 AAGGTGAAGACCCCTCATCGAGACCGCGCATAGTCGCCGCCCAAGGAGATAGCGGAC 1263
 297 LeuGlyTyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsn 316
 1264 ATCGCTTAC-----GGTGGGTAGGTGAACGCTCTACAGGCC----- 1302
 317 ValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGln-----Lys 331
 1303 ATCAAGTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1362
 332 AlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValThrThr 351
 1363 GCCACCCACACCTTCGAGCTCAGCGCGGCGCACCTTCGACCGCCACCTCTACTGGGAC 1422
 352 AspAlaProGlySerThrAlaSerTyrThrLeuValAsnAspLeuValIle 371
 1423 -----ACGGGCTCGAGGAGTGAACGCTCTACCTTACCTTACCTTACCTTACCTTAC 1452
 372 ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsn 391
 1453 TAGACCCCAAC-----GGGACGAGGTGACTTACTTACTTACTTACTTACTTACTTACTT 1497
 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyr 411
 1498 TAC-----GGCTTCGAGAGTGGCTACTTACTTACTTACTTACTTACTTACTTACTT 1545
 412 ThrIleGluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
 1546 ACGGTCAAGTCTCGTCACTAC-----AAGGGCGCGCGGAACTTACCAAGTCTCGACGTC 1596
 432 Val 432
 1597 GTC 1599

RESULT 6
 US-09-445-472-11
 ; Sequence 11, Application US/09445472
 ; Patent No. 6358726
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: SHIMOJO, Tomoko

APPLICANT: ASADA, Kiyozo
 APPLICANT: KATO, Kunoshin
 TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
 FILE REFERENCE: TAKAKURA=6
 CURRENT APPLICATION NUMBER: US/09/445,472
 CURRENT FILING DATE: 1999-12-06
 PRIOR APPLICATION NUMBER: 151969/1997
 PRIOR FILING DATE: 1997-06-10
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: Patent in version 3.0
 SEQ ID NO 11
 LENGTH: 1977
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE: Information: Synthetic
 OTHER INFORMATION: US-09-445-472-11

Alignment Scores:
 Pred. No.: 1,4e-34 Length: 1977
 Score: 443.00 Matches: 137
 Percent Similarity: 46.03% Conservative: 66
 Best Local Similarity: 31.07% Mismatches: 170
 Query Match: 19.59% Indels: 68
 Gaps: 16

US-09-985-689A-5 (1-433) x US-09-445-472-11 (1-1977)

QY 8 ValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnGlnValValAla 27
 Db 433 ATAGGGCCCATACCGTCTCGAACTCCCTCGGTACAGCGAAGCGGTGGTGGTGGC 492
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArg 47
 Db 493 ATCGTCGATACGGGTATAGACGGGAAC-----CACCCCGATCTGAAG 534
 QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnProAsn 66
 Db 535 GGCAGGTCTACAGCTGGTACGACGCGCTCAAGCGAGGTCGACCCCTACGATGACCAG 594
 QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
 Db 595 GGACACGGACCCACGTTGCGGTATGTTGCCGGAACCGGACGCGTAACTCCAGTAC 654
 QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 100
 Db 655 ATAGCGCTCGCCCGCGCGGAGCTCTCGCGGTCAAGGTCTCTCGGTGCGCGGTTCG 714
 QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAla 120
 Db 715 GGAAGCTCTCCACCATCATCGCGGTGTTGACTGGCTCGTCCAGAAACAAGGACAGTAC 774
 QY 121 GlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
 Db 775 GGGATAAGGTTCATCAACCTCTCCCTCGGTCTCTCCAGAGCTCCGAGAACCGGACTCC 834
 QY 141 SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAagly 160
 Db 835 CTCAGTCAGCGGCTCAACACCCCTGGACCGCGGTATAGTAGTCTCGTCCGCCCGGC 894
 QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 Db 895 AACAGCGGGCGGAACACCTACAGCTCGGTCTCACCCCGCGCGGACCAAGGTGCATAACC 954
 QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
 Db 955 GTCGGTGCA-----GTTGACAGCAACGACAC 981
 QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 Db 982 ATCGCCAGCTTCTCCAGCAGGGGACCGACCGCGGACCGAAGCTCAAGCCGGAAGTCTGTC 1041
 QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240

```
Db 1042 CCCCCCGCGTTGACATCATAGCCCGCGCCAGC-----GGAACCGCATGGC 1092
Qy 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
Db 1093 ACCCGATAACGACTACTACACAGCCCTCTGGACCGCATGGCCACCGCGAGTT 1152
Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro--- 279
Db 1153 TCGGCGGTGGCGGCTCATCTCCAG-----GCCACCGCGAGTGGACCCCGGAC 1203
Qy 280 -----LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly--- 296
Db 1204 AAGTGAAAGACCGCCCTCATCGACCCCGACATAGTCGCCCCCAAGGATAGCGGAC 1263
Qy 297 LeuGlyTyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsn 316
Db 1264 ATCGCTAC-----GGTGGCGTAGGTGAACGTCTACAGGCC----- 1302
Qy 317 ValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGln-----Lys 331
Db 1303 ATCAAGTACGACGACTAGCAGCTACCTTACCGGCTCCGTCGCCGACACAGGAGC 1362
Qy 332 AlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValTyrThr 351
Db 1363 GCCACCCACACCTTCGACGTCAGCGGCCACCTTCGTGACCGCCACCTCTACTGGAC 1422
Qy 352 AspAlaProGlySerThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIle 371
Db 1423 -----ACGGGCTCGAGCGACATCGACCTCTACCTC 1452
Qy 372 ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsn 391
Db 1453 TAGCACCCCAAC-----GGGAACGAGGTGTACTCTCCTACACGCCCTAC 1497
Qy 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyr 411
Db 1498 TAC-----GGCTTCGAGAAGGTGGCTACTACAACCCCGACCGCGGACCTGG 1545
Qy 412 ThrIleGluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
Db 1546 ACGGTCAAGTGTGTAGCTAC-----AAGGGCGGCGGCAACTACCAGGTGCGACGTC 1596
Qy 432 Val 432
Db 1597 GTC 1599
```

RESULT 7

```
US-09-445-472-2
; Sequence 2, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-2
```

Alignment Scores:

```
Pred. No.: 7.65e-33 Length: 1236
Score: 422.50 Matches: 138
Percent Similarity: 43.96% Conservative: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.69% Indels: 79
DB: 4 Gaps: 17

US-09-985-689A-5 (1-433) x US-09-445-472-2 (1-1236)

Qy 16 AsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 58 AACTGGGATATGATGTTCTCGAATCAATAGGAATAATTCACACTGGAATTGAC--- 114
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 115 -----GCITTCATCCAGATCTCCAGGAAAGTA----- 144
Qy 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 145 ATTGGGTGGGTAGATTCTTCAATGGTAGGAGTTATCCATACGATGACCATGACATGGA 204
Qy 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db 205 ACTCATGTAGCTTCAATAGCAGCTGTACTGGAGCAGCAAGTAATGGCAAGTACAGGA 264
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db 265 ATGCTCCAGGAGCTAAGCTGGCGGAATTAAGTTCTAGGTCCGCGATGTTCTGGAAGC 324
Qy 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
Db 325 ATATCTACTATATATTAAAGGAGTTAGTGGCGCGCTTGATAACAAGATAGTACGGAAT 384
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
Db 385 AAGTCATTAATCTTCTCTGTTCAAGCCAGAGCTCAGATGGTACTCAGCGTCTTAAGT 444
Qy 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaGlyAsnGlu 162
Db 445 CAGGCTGTTAATGACGCGTGGGATGCTGATAGTTGTTGTTGCTGCCCTGGAACAGT 504
Qy 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 505 GGACCTTAACAAGTATACATCGGTTCTCAGCAGCTGCAAGCAAACTTATTACAGTTGA 564
Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla 202
Db 565 GCGGTTGACAAAGTAT-----GATGTTATAACA 591
Qy 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 592 AGCTTCTCAAGCAGAGGCGCAACTGCGAGCGGCGCTTAAGCTGAGGTGTTGTTCTCCA 651
Qy 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 652 GGAACCTGGATAATTGCTGCCAGAGCAAGT-----GGAACCTAGCATGGGTCAACCA 702
Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrAspValGly--- 262
Db 703 ATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCAACTCTCTCAGTAGCTGT 762
Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
Db 763 ATTGCAGCCCTCTGCTCCAA-----GCACCCCGAGCTGGACTCCAGACAAAGTA 813
Qy 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
Db 814 AAAACAGCCCTCATAGAAACTGCTGATATCGTAAGCCAGATGAAATAGCCGATATAGCC 873
Qy 299 TyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAla 318
Db 874 TAC-----GGTGCAGGTAGGTTAATGCATACAGCGCTATAAAC----- 912
```

QY 319 TyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAla-----Thr 333
Db 913 TACGATAACTATGCAAGCTAGTGTCTACTGATATGTCACAAAGAGGCGCAAACT 972
QY 334 TyrSerPheGlnAlaGlnAlaGlyProLeuLysIleSerLeuValThrAspAla 353
Db 973 CACCAAGTTCGTTATTAGCGAGCTTCGTTGTAACCTGCCACATTATCTGGCAATGCC 1032
QY 354 ProGlySerThrThrAlaSerThrLeuValAsnAspLeuAspLeuValIleThrAla 373
Db 1033 AAT-----ACGACCTTGATCTTACTCTACGAT 1062
QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnThrAsp 393
Db 1063 CCCAATGGAACACAGCTT-----GACTACTCTTAC-----ACC 1095
QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrIle 413
Db 1096 GCTACTATGAGTTCGAAAGGTTGTTATTACCAACCCAACTGATGGACATGGACAATT 1155
QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1156 AAGGTTGTAAGCTAC-----ACCGAAGTGCACAACTATCAAGTAGATGGTA 1203

RESULT 8

US-08-894-818B-4
; Sequence 4, Application US/08984818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; OTHER INFORMATION: /note= N at position 1283 is G or T.
US-08-894-818B-4

Alignment Scores:
Pred. No.: 11e-32 Length: 1566
Score: 422.50 Matches: 138
Percent Similarity: 43.96% Conservative: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.69% Indels: 79
Dbs: 17

US-09-985-689A-5 (1-433) x US-08-894-818B-4 (1-1566)

QY 16 AsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 58 AACTTGGGATATGATGTTCTGGAATCAACAATAGGAATAATTGACATCGAATTGAC--- 114
QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 115 -----GCTTCTCATCCAGATCTCCAGGAAAGTA----- 144
QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 145 ATTGGGTGGGTAGATTTTGTCAATGTTAGGAGTTATCCATACGATGACCATGACATGGA 204
QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db 205 ACTCATGTAGCTTCAATAGCAGCTGTACTGGAGCAGCAAGTAATGGCAAGTACAGGGA 264
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db 265 ATGGCTCAGGAGCTTAAGCTGGCGGAATTAAGGTTCTAGGTCCGATGTTCTTGGAGC 324
QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAlaGlyAla 122
Db 325 ATATCTACTATATTAAAGGAGTTAGTGGCGCTTGATACAAAGATAAGTACGGAATT 384
QY 123 ArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
Db 385 AAGGTCATTAATCTTCTCTGTTTCAAGCCAGAGCTCAGATGTTACTGACGCTCTAAGT 444
QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 445 CAGGCTGTAAATGACGCGTGGGATGCTGGATTAGTTGTTGTTGCTGCCCTGGAAACAGT 504
QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 505 GGACCTAACAGATATACATCGGTTCTCCAGCAGCTGCAAGCAAAAGTTATTACAGTTGGA 564
QY 183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla 202
Db 565 GCCGTTGCAAGTAT-----GATGTTATTACA 591
QY 203 GlnPheSerSerArgGlyValaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 592 AGCTTCTCAGCAGAGGCGCAACTCGACGCGCAGGCTTAAGCTAGGTTGTTGCTCCA 651
QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsn 242
Db 652 GGAAACTGGATAATTGCTGCCAGAGCAAGT-----GGAACTAGCATGGGTCAACCA 702
QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 703 ATTAATGACTATTACACAGAGCTCCTGGGACATCAATGGCACTCTCTCAGTAGTGGT 762
QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
Db 763 ATTGCAGCCCTCTTGTCTCAA-----GCACACCCGAGCTGGACTCCAGACAAAGTA 813
QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly----LeuGly 298

[illegible]

QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
Db 1459 CCAATGGAAACAGGTT-----GACTACTCTTAC-----ACC 1491
QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrile 413
Db 1492 GCTACTATGAGTTCGAAAGGTTGGTTATTACACCCAACTGATGGAAACATGGCAATT 1551
QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1552 AAGTTGTAAAGTAC-----ACCGGAAGTGCACAACTATCAAGTAGATGTGTA 1599

RESULT 11

US-08-894-818B-6
; Sequence 6, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESS: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-894-818B-6

Alignment Scores:
Pred. No.: 1.99e-30
Score: 401.50
Percent Similarity: 43.96%
Best Local Similarity: 29.61%
Query Match: 17.76%
DB: 3

US-09-985-689A-5 (1-433) x US-08-894-818B-6 (1-1977)

QY 16 AsnTyrGlyLeuTyrGlyGlnGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 454 AACTTGGCATATGATGTTCTTGAATCACAATAGGAATAATGACACTGGAATTGAC--- 510
QY 36 GlyArgAsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 511 -----GTTCTCATCCAGATCTCCAGGAAGAAGTA----- 540
QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 541 ATTGGTGGGTAGATTTTGTCAATGGTAGGAGTTATCCATACGATGACCATGACATGGA 600
QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db 601 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCACAAGTATATGGCAAGTACAGGGA 660
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db 661 ATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGGTTCTAGGTGCCGATGTTCTTGAAGC 720
QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAlaGlyAla 122
Db 721 ATATCTACTATTAATTAAGGGAGTTGAGTGGCCCTTGATACAAAGATAAGTAGCGGAAT 780
QY 123 ArgIleHisThrAsnSerTrpGlyValaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
Db 781 AAGTCTAATTAATCTTCTTGTGTTCAAGCCAGAGCTCCGAGCGAACCGACTCCCTCAGT 840
QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 841 CAGGCGGTCAACAACGCTGGGACGCGGTATAGTCTGCTGCGCGCGCGGCAACAGC 900
QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 901 GGGCGGAACACTACACGCTGGCTCACCCTCCCGCGGAGCAAGGTACATACCGTCGGT 960
QY 183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla 202
Db 961 GCA-----GTTGACACGACCAACATCGCC 987
QY 203 GlnPheSerSerArgGlyValaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 988 AGCTTCTCAGCGGGGACCGACCGCGGACGAGAGGTCAAGCCGGAAGTGTGTCGCCCC 1047
QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1048 GCGGTTGACATCATAGCCCGCGCGCCAGC-----GGAACCGACATGGGCACCCCG 1098
QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1099 ATAAACGACTACTACCAAGGCTCTTGGAAACAGCATGGCCACCCCGACGTTTCGGGC 1158
QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
Db 1159 GTTGGCGGCTCATCTCCAG-----GCCACCGAGCTGGACCCCGCAAGGTG 1209
QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
Db 1210 AAGACCGCCTCATCGACCGCGGACATATGCGCCCCCAAGGAGATAGCGACATCGCC 1269
QY 299 TyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 318
Db 1270 TAC-----GGTGGGTAGGTGAACGCTACAGGCC-----ATCAAG 1308
QY 319 TyrValAsnGluAlaThrAlaLeuAlaThrGlyGln-----LysAlaThr 333
Db 1309 TACGACGACTAGCCCAAGCTCACCTTCCACCGCTCCGTCGCGCAAGGAGCGCCACC 1368
QY 334 TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValThrThrAspAla 353
Db 1369 CACACCTTCGACGTCAGCGGGCGCCACCTTCGTGACGCGCCACCTCTACTTGGAC----- 1422

QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAlaAspLeuAspLeuValIleThrAla 373
Db 1423 -----ACGGCTCGAGCGACATCGACCTTACCTTACGAC 1459
QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
Db 1459 CCAAC-----GGACAGAGTTGACTACTCTACACCGCTACTAC--- 1500
QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413
Db 1501 -----GGCTTCGAGAGTTCGGCTACTACACCGACCGCGGAACCTGGACGGTC 1551
QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1552 AAGTCTCGACTAC-----AAGGCGCGCGGAACCTACAGTCTCGACTCGTC 1599

RESULT 12
US-09-000-016-3
; Sequence 3, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: AKIRA ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2539 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces viridosporus
; STRAIN: A-914
; FEATURES:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
US-09-000-016-3

Alignment Scores: 3 7e-25 Length: 2539
Pred. No.: 350.50 Matches: 134
Score:

Percent Similarity: 41.28% Conservative: 53
Best Local Similarity: 29.58% Mismatches: 168
Query Match: 15.50% Indels: 98
DB: 3 Gaps: 19

US-09-985-689A-5 (1-433) x US-09-000-016-3 (1-2539)

QY 18 GlyLeuTyrGlyGlnGlnValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 998 GGCTACGACGGCAAGGCGGTGAAGATCGCGTCTGTGACACCGGTGTGACACGAGC--- 1054
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 1055 -----CATCCGGACCTGAGAGCGGGGTGACCGGTCCTCAAGAACTTCACC 1099
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1100 GCCCGCGCGCGCGCGGACAGGTGGCGCACAGGTGGCGCACCGCACCTCGCTCGCGGCG 1159
QY 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 1160 GGCACGCGCGCGCCAGTCCAAGGGCAAGTACAAGGGCGGTGCGACCGCGCGCGGATCCTC 1219
QY 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThr 112
Db 1220 AACGCAAGTCTCTCGACGACTCC-----GGTTTCGCGCAGCAGCTCCGGCATC 1267
QY 113 LeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHisThrAsnSerTrpGly 130
Db 1268 CTCGCGGCATGAGTGGCG 1327
QY 131 Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
Db 1328 GGCATGGACACACCGGAGACCGCGCTGAGGCGGCG----- 1366
QY 144 ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly 163
Db 1367 GTCGCAAGTGTCTCGCGGAGAGGGCGTCTCTGTTCGCATCGCGCGCGCGCAAGAGGCG 1426
QY 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 1427 CCGGAGTCG-----ATCGTTCGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCG 1480
QY 184 ThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAlaGln 203
Db 1481 -----GTCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1507
QY 204 PheSerSerArgGly-----AlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 1508 TTCTCTCTCCACCG 1567
QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1568 GCGGTGGACATCACG 1627
QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1628 GGACCGCGCGCGCTACATGACCATCTCCGCGACGTCGCGCGCGCGCGCGCGCGCGCGCG 1687
QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSer 282
Db 1688 GCGCGCGCGCTCTGAGCAGCAG----- 1720
QY 283 LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1721 TGGACCTCGCGCGCACTGAAGGCGCGCGCTCACCGGCTCCACCAAGGGCGCGC--- 1777
QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1778 ACCCGGTCGAGCAGCGGTTCGGCGCGGATCCAGCGCGACAGCGCGTCCAGCAGACCGGTG 1837
QY 320 ValAsnGluAlaThrAlaLeuAlaThrGly----- 329


```
QY 204 PheSerSerArgGly---AlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db |||||
Db 1508 TTCTCTCCACCGCGCCCGCTCGCGACGGCCATCAAGCCGCGCTCACCGCTCCC 1567
QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db |||||
Db 1568 GCGTGGACATCACCGCGCTCGCGGAGGCGACGACATCGCGGAGGTCGGTGAG 1627
QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db |||||
Db 1628 GGACCGCGCGGTACATGACCATCTCCGCGACGTGATGGGACCCCGACGTCGCGGCG 1687
QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSer 282
Db |||||
Db 1698 GCGCGCGCTCTGAGCAGCAG-----CACCCCGAC 1720
QY 283 LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db |||||
Db 1721 TGGACCTCGCGGAATCAAGCGCGCTCACCGCTCCACCAAGGCGGC---AAGTAC 1777
QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db |||||
Db 1778 ACCCGTTCGACGAGGTCGCGCGCGATCCAGCGCGACAGCGCTCCAGCAGCGGTG 1837
QY 320 ValAsnGluAlaThrAlaLeuAlaThrGly----- 329
Db |||||
Db 1838 ATCGCGACCGCTCTCGGTGAGTTCGCGCTCCAGCAGTGCAGCACCGACGACGAG 1897
QY 330 -----GlnIysAlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuIysIle 346
Db |||||
Db 1898 CCGGTACCAAGCAGCTGACCTACCGACCTCGCGACCCAGCGTCAAGTCAAGTGT 1957
QY 347 SerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----TyrThrLeuVal 364
Db |||||
Db 1958 ACGTCACCGCACCGACCCCAAGCGCGCGCGCGCGCTTCTTACGCTGGGC 2017
QY 365 AsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPhe 384
Db |||||
Db 2018 -----GCCACCAAGTGCACCGTCCGCGCGCGCGC----- 2047
QY 385 SerTyrProTyrAspAsnAsnTrpAspGlyArg-----AsnAsnValGlu 399
Db |||||
Db 2048 AGCGCTCGTGCATGACGATGCCCGCGACCGCGCGCGCGCGCGTGGCGCGGTAC 2107
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle----- 413
Db |||||
Db 2108 TCGCGTACGTGTCGCC---ACGGCGCGCGCGCGACGCTCCGCGCGCGCGCGCGGTG 2164
QY 414 -----GluValGlnAlaTyrAsnVal-ProSerGly 423
Db |||||
Db 2165 CAGCGCGAGTTCAGTGTGACGACGTGACCGTCCGCGCGCGC 2201

RESULT 15
US-09-514-340-1
; Sequence 1, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Wendroth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
```

```
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA: US/09/514,340
APPLICATION NUMBER: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN: <Unknown>
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-514-340-1

Alignment Scores:
Pred. No.: 1.37e-24 Length: 2809
Score: 345.50 Matches: 133
Percent Similarity: 41.06% Conservative: 53
Best Local Similarity: 29.36% Mismatches: 169
Query Match: 15.28% Indels: 98
DB: 4 Gaps: 19

US-09-985-689A-5 (1-433) x US-09-514-340-1 (1-2809)
QY 18 GlyLeuTyrGlyGlnGlyGlnValAlaValAlaAlaAspThrGlyLeuAspThrGlyArg 37
Db |||||
Db 998 GGCTACGACGCGCAAGGCGGTGAAGATCGCGCTCTCGACACCGGTGTGACACGAGC--- 1054
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db |||||
Db 1055 -----CATCGGACCTGAAGGCGGTGACCGGTGCCGTCACGAAGACTTCACC 1099
QY 58 ArgThrAsnAlaAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db |||||
Db 1100 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1159
QY 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db |||||
Db 1160 GGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 1219
QY 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThr 112
Db |||||
Db 1220 AACGCGAAGTCTCTCGACGACTCC-----GGTTTCGCGACGACTCCGCGCATC 1267
QY 113 LeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHisThrAsnSerTrpGly 130
Db |||||
Db 1268 CTCGCGCGCATGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1327
QY 131 Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
Db |||||
```

1328 GGCATGGACACACCGGAGACCGACCCCGCTGGAGGGGGG----- 1366
144 ValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly 163
1367 GTCGACAAAGCTGCCCGGAGAGGGGCTCTGTCGCCATCGCGCGGCAACGAGGGC 1426
164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAlaIleThrValGlyAla 183
1427 CCGAGTCG-----ATCGGTTCCCGCGGAGCGGCGCCCTCACCGTCGCGGCC 1480
184 ThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAlaGln 203
1481 -----GTCGACGACAAAGCAAGCTCGCGGAC 1507
204 PheSerSerArgGly---AlaThrArgAspGlyArgIleIysProAspValThrAlaPro 222
1508 TTCTCTCCACCGCGCCCGCTCGCGGACGCGGCCCATCAAGCGGACGTCAACCGCTCCC 1567
223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
1568 GCGGTGGACATCACGGCGGCTCGGGGAGGGCAACGACATCGCGCGGAGGTCGGTGAG 1627
243 TyrAsnSerIysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 262
1628 GGACCGCGCGCTACATGACCATCTCCGGCACGTCGATGGCGACCCCGCACGTCGCGGGC 1687
263 AsnValAlaGlnLeuArgGluHisPheIleIysAsnArgGlyIleThrProLysProSer 282
1688 GCGCGCGCCCTCTGAAGACGAG-----CACCCCGAC 1720
283 LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
1721 TGGACCTCCGCGCAACTGAGGGCGCGCTCACCGGTCCACCAAGGGCGGC---AAGTAC 1777
300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
1778 ACCCGTTCGAGCAGGGTTCGGCGCGATCCAGGCCACAGCGCTCCAGCAGACCGTG 1837
320 ValAsnGluAlaThrAlaLeuAlaThrGly----- 329
1838 ATCGCCGACCCGGTCTCGGTGAGTCTCGGCTCCAGCAGTGGCGGCACACCGACGACGAG 1897
330 -----GlnLysAlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIle 346
1898 CCGGTACCAAGCAGCTGACCTACCGCAACTCGGCACCCAGGACGTCACGCTGAAGCTG 1957
347 SerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----TyrThrLeuVal 364
1958 ACGTCACCGCCACCGACCCCAAGGCAAGCGGCGCGGCTTCTTCAAGCTGGGC 2017
365 AsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPhe 384
2018 -----GCCACACCGGTACCGTCCCGCGGGCGGC----- 2047
385 SerTyrProTyrAspAsnAsnTrpAspGlyArg-----AsnAsnValGlu 399
2048 ACGGCTCGTCGACATACCGCCGACACCGGCTCGCGGCGACGCTGGACGGCGCGTAC 2107
400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle----- 413
2108 TCGGGGTACGTGTCGCC---ACGGCGCGGGGACGAGCGTCCGACGCGCGCGCGGTG 2164
414 -----GluValGlnAlaTyrAsnVal-ProSerGly 423
2165 CAGCGCGAGTCTGAGTCTGATGACGTGACCGTCCCGGC 2201

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2004, 08:10:21 ; Search time 327.248 Seconds
(without alignments)
4948.852 Million cell updates/sec

Title: US-09-985-689A-5
Perfect score: 2261
Sequence: 1 NDVARGIVKADVAQNNGY.....EVQANVPSGQRFSLAIVH 433

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2466186 seqs, 1870095128 residues

Total number of hits satisfying chosen parameters: 4932372

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US0985689/runat_31032004_161809_4271/app_query.fasta_1.3498
-DB=Published Applications NA -QWTF=fastap -SUPPLX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORW=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US0985689 @CGN 1 1 601 @runat_31032004_161809_4271
-NCPUS=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	1998.5	88.4	1305	15	US-10-385-662-1	Sequence 1, Appli
2	443	19.6	1977	13	US-10-090-624-11	Sequence 11, Appl
3	422.5	18.7	1236	13	US-10-090-624-2	Sequence 2, Appli
4	422.5	18.7	1962	13	US-10-090-624-15	Sequence 15, Appl
5	341.5	15.1	135638	14	US-10-314-657-1	Sequence 1, Appli
6	334	14.8	1329	9	US-09-974-300-1934	Sequence 1934, Ap
7	329	14.6	3417	14	US-10-156-761-3306	Sequence 3306, Ap
8	329	14.6	9025608	14	US-10-156-761-1	Sequence 1, Appli
9	327	14.5	3624	14	US-10-156-761-5701	Sequence 5701, Ap
10	327	14.5	9025608	14	US-10-156-761-1	Sequence 1, Appli
11	307	13.6	4785	13	US-10-090-624-5	Sequence 5, Appli
12	298.5	13.2	3743	10	US-09-927-827-29	Sequence 29, Appl
13	295	13.0	1560	15	US-10-084-846A-113	Sequence 113, App
14	295	13.0	59816	15	US-10-084-846A-2	Sequence 2, Appli
15	295	13.0	59816	15	US-10-084-846A-2	Sequence 2, Appli
16	284.5	12.6	3303	14	US-10-156-761-5384	Sequence 5384, Ap
17	276.5	12.2	2166	12	US-10-344-231-17	Sequence 17, Appl
18	276.5	12.2	2166	12	US-10-363-332A-17	Sequence 17, Appl
19	273	12.1	3788	10	US-09-927-827-33	Sequence 33, Appl
20	269	11.9	1140	8	US-08-322-678-11	Sequence 11, Appl
21	269	11.9	1140	16	US-10-323-324-11	Sequence 11, Appl
22	269	11.9	1143	14	US-10-313-853-6	Sequence 6, Appli
23	269	11.9	2588	12	US-10-344-231-20	Sequence 20, Appl
24	269	11.9	2588	12	US-10-363-332A-20	Sequence 20, Appl
25	268	11.9	1140	8	US-08-322-678-12	Sequence 12, Appl
26	268	11.9	1140	16	US-10-323-324-12	Sequence 12, Appl
27	260.5	11.5	1306	9	US-09-966-921A-1	Sequence 1, Appli
28	260.5	11.5	1330	9	US-09-966-921A-5	Sequence 5, Appli
29	259.5	11.5	840	14	US-10-209-812-1	Sequence 1, Appli
30	258.5	11.4	2267	12	US-10-344-231-18	Sequence 18, Appl
31	258.5	11.4	2267	12	US-10-363-332A-18	Sequence 18, Appl
32	252	11.1	1485	9	US-09-974-300-1938	Sequence 1938, Ap
33	251	11.1	3884	10	US-09-927-827-33	Sequence 34, Appl
34	250.5	11.1	1971	9	US-09-974-300-1935	Sequence 1935, Ap
35	247.5	10.9	2192	12	US-10-424-599-112429	Sequence 112429,
36	242	10.7	1140	15	US-10-146-905A-9	Sequence 9, Appli
37	240.5	10.6	3452	10	US-09-927-827-30	Sequence 30, Appl
38	238	10.5	1140	9	US-09-920-118-13	Sequence 13, Appl
39	231.5	10.2	4338	9	US-09-891-711-3	Sequence 3, Appli
40	231.5	10.2	4338	12	US-10-432-887-953	Sequence 953, App
41	230	10.2	1332	14	US-10-156-761-5689	Sequence 5689, Ap
42	230	10.2	1575	12	US-10-344-231-19	Sequence 19, Appl
43	230	10.2	1575	12	US-10-363-332A-19	Sequence 19, Appl
44	228	10.1	4198	9	US-09-891-711-5	Sequence 5, Appli
45	227.5	10.1	522	10	US-09-824-893A-98	Sequence 98, Appl

ALIGNMENTS

RESULT 1
US-10-385-662-1
; Sequence 1, Application US/10385662
; Publication No. US20040002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAITO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAEKI, KATSURISA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938USO
; CURRENT APPLICATION NUMBER: US/10385,662
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231

US-09-985-689A-5 (1-433) x US-10-090-624-11 (1-1977)

QY 8 ValLysAlaAspValAlaGlnAsnTyrGlyLeuTyrGlyGlnValValAla 27
DB 433 ATAGGGCGGATACCGTCTGGAACCTCCCTCGCTACGACGGAAGCGGTGTGGTGGTCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArg 47
DB 493 ATCGTCGATACGGGTATAGACGGNAC-----CACCCGATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
DB 535 GGCAGGTATAGGTGTGTAGACCGCTCAACGCGAGTCCGACCCCTACGATGACGAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
DB 595 GGACACGGAACCCAGCTTGGCGGTATCGTTGCCGGAACCGGACGGTGTAACTCCAGTAC 654
QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 100
DB 655 ATAGCGTCTGCGCGCGCGGCGAAGCTCTCGCGGTCAAGGTTCTCGGTGCGGACGGTTCG 714
QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaThrPAsnAla 120
DB 715 GGAACGCTCTCCACCATCATCGCGGTGTGACTGGTCCGACGAAACGAGCAAGTAC 774
QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
DB 775 GGGATAAGGTATCAACCTCTCCCTCGCTCCCTCCAGAGCTCCGACGGAACCGACTCC 834
QY 141 SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
DB 835 CTCAGTCAGCGCGTCAACACACCCCTGGGACGCGGTATAGTAGTCTGCGTCCGCGCGGC 894
QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
DB 895 ACACGCGCGCGAACAACCTACACCTCGCTCACCCTCGCGCGGACGAGCAAGTATAC 954
QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
DB 955 GTGCGTGCAC-----GTTGACGAGCAACGACCAAC 981
QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
DB 982 ATCGCAGCTTCTCCAGCGGAGCGACCGCGGAGCGAAGGCTCAACCGGAGTCTGTC 1041
QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
DB 1042 GCGCGCGGTGGATCATCATGCGCGCGCGCAGC-----GGAACACGATGCGC 1092
QY 241 AlaAsnTyrAsnSerLysThrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
DB 1093 ACCCGATTAACGACTACTACACAGGCTCTGGAACGAGCATGGCCACCGCGACGTT 1152
QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro--- 279
DB 1153 TCGGGCGTGGCGGTCTCTCCAG-----GCCACCGCGAGTGGACCGCGGAC 1203
QY 280 -----LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly--- 296
DB 1204 AAGGTGAAGACCGCTCATGACACCGCGCATAGTCCGCCCAAGGAGATACGAG 1263
QY 297 LeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsn 316
DB 1264 ATCGCTAC-----GGTGGGTAGGTGAACGTCTACAGGCC----- 1302
QY 317 ValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGln-----Lys 331
DB 1303 ATCAAGTACGACGACTACGCAAGCTCACCTTCACCGCTCGCTCGCGCAAGGAAGC 1362
QY 332 AlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValThr 351

DB 1363 GGCACCCACACCTTCGACGTCAGCGCGCCACCTTGTGTGACCGCCACCTCTACTGGGAC 1422
QY 352 AspAlaProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValle 371
DB 1423 -----ACGGGCTCGAGGACATCGACCTCTACTCTTACCTC 1452
QY 372 ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsn 391
DB 1453 TAGCAGCCCAAC-----GGGAACGAGGTGTACTCTCTACACCGCCTAC 1497
QY 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyr 411
DB 1498 TAC-----GGCTTCGAGAGGTGCGCTACTACACCCGACCGCCGGAACCTGG 1545
QY 412 ThrIleGluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
DB 1546 ACGTCAAGGTCTGTCAGTAC-----AAGGGCGCGGCACTACCAGGTGCGAGTC 1596
QY 432 Val 432
DB 1597 GTC 1599

RESULT 3
US-10-090-624-2
; Sequence 2, Application US/10090624
; Publication No. US2002012335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR FILING DATE: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR FILING DATE: 151969/1997
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-2

Alignment Scores:
Pred. No.: 1,76e-36 Length: 1236
Score: 422.50 Matches: 138
Percent Similarity: 43.96% Conservative: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.69% Indels: 79
DB: 13 Gaps: 17

US-09-985-689A-5 (1-433) x US-10-090-624-2 (1-1236)

QY 16 AsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
DB 58 AACTTGGGATATGATGGTCTCGAATCAACATAGGAATTAATTCACACTGGAATTGAC--- 114
QY 36 GlyArgAsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
DB 115 -----GCTTCTCATCCAGATCTCCAGAAAGTA----- 144
QY 56 LeuGlyArgThrAsnAsnAlaAla-----AspProAsnGlyHisGly 69
DB 145 ATTGGGTGGGTAGATTGTCATGATGGTAGGAGTTATCCATGATGACCATGACATGGA 204
QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84

Db 205 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAAGTACAGGGA 264
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db 265 ATGGCTCCAGGAGCTAAGCTGGGGGAATTAAGTTCTAGGTGCCGATGGTCTGGAAGC 324
Qy 103 LeuGlyClyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
Db 325 ATATCTACTATAATTAAGGAGTTGAGTGGCGGTGGATTAACAAGATAAGTACGGAAT 384
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaValTrpAlaAsnSerArg 142
Db 385 AAGTCAATTAATCTTCTCTGTTCAAGCCAGAGCTCAGATGGTACTGACGCTCTAAGT 444
Qy 143 GlnValAspGluTrpValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 445 CAGGCTGTTAATGACGCTGGGATGCTGGATTAGTTGTTGTCGCCCTGGAAACAGT 504
Qy 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 505 GGCCTTAACAAGTATACATCGTTCTCCAGCAGCTGCCAAGCAAGTTATTACAGTTGA 564
Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla 202
Db 565 GCGGTTGACAAGTAT-----GATGTTATAACA 591
Qy 203 GlnPheSerSerArgGlyAlaThrArgAspGlyValGlyLeuLysProAspValThrAlaPro 222
Db 592 AGTTCTCAAGCAGAGGGCCCAATGCGACGCGGAGCTTAAGCTGAGGTGTTGTTCTCA 651
Qy 223 GlyThrPheLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 652 GGAACCTGGATAATGTCGACAGCAAGT-----GGAACCTAGCAGTGGTCAACCA 702
Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 703 ATTAATGACTATTACACAGCAGCTCTCGGACATCAATGGCAACTCTCAGTGTGTTGT 762
Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
Db 763 ATTCAGCCCTCTTGCTCCAA-----GCACACCGAGCTGGACTCCAGACAAGTA 813
Qy 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
Db 814 AAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAATAGCCGATATAGCC 873
Qy 299 TyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 318
Db 874 TAC-----GGTGCAGGTAGGTTAATGCATACAGGCTATAAC----- 912
Qy 319 TyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAla-----Thr 333
Db 913 TACGATAACTATGCAAAAGCTAGTTCTCAGTGGATATGTTGCCAACAAAGCGCAACCACT 972
Qy 334 TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAla 353
Db 973 CACCAGTTCGTATTAGCGGAGCTCGTTCGTAACTGCCATATTAAGTGGCAATGCC 1032
Qy 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuValIleThrAla 373
Db 1033 AAT-----AGCGACCTTGATCTTTTACCTACGAT 1062
Qy 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
Db 1063 CCCAATGGAACCAAGTT-----GACTACTCTTAC-----ACC 1095
Qy 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrIle 413
Db 1096 GCCTACTATGATTGCAAAAGTTGGTTATTACAAACCAACTGATGGAACATGGACAAT 1155
Qy 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432

Db 1156 AAGTTTGAAGCTAC-----AGCGAAGTGCAAACTATCAAGTAGATGGTA 1203
RESULT 4
US-10-090-624-15
; Sequence 15, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-15
Alignment Scores:
Pred. No.: 3,45e-36 Length: 1962
Score: 422.50 Matches: 138
Percent Similarity: 43.96% Conservative: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.69% Indels: 79
DB: 13 Gaps: 17
US-09-985-689A-5 (1-433) x US-10-090-624-15 (1-1962)
Qy 16 AsnTyrGlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 454 AACTGGGATATGATGTTCTGGAATCAATAGCAATAATAGACATGGAATTGAC--- 510
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 511 -----GCTTCTCCAGATCTCCAGGAAGAATA----- 540
Qy 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 541 ATTGGTGGGTAGATTTTGTCAATGGTAGGAGTTATCCATACGATGACCATGGACATGGA 600
Qy 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db 601 ACTCATGTAGTTCATATAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAAGTACAGGGA 660
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db 661 ATGGCTCCAGGAGCTAAGCTGCGGGAATTAAGTTCTAGGTGCCGATGGTCTGGAAGC 720
Qy 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
Db 721 ATATCTACTATAATTAAGGAGTTGAGTGGCGCTTGATAACAAGATAAGTACGGAAT 780
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaValTrpAlaAsnSerArg 142
Db 781 AAGTCAATTAATCTTCTCTGTTCAAGCCAGAGCTCAGATGGTACTGACGCTCTAAGT 840
Qy 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 841 CAGGCTGTTAATGACGAGCTGGGATGCTGATAGTTGTTGTTGTCGCCCTGGAAACAGT 900
Qy 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182

```

; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 135638
; TYPE: DNA
; ORGANISM: Streptomyces atroolivaceus
US-10-314-657-1

Alignment Scores:
Pred. No.: 1,316-24 Length: 135638
Score: 341.50 Matches: 139
Percent Similarity: 42.17% Conservative: 55
Best Local Similarity: 30.22% Mismatches: 155
Query Match: 15,10% Indels: 111
DB: 14 Gaps: 23

US-09-985-689A-5 (1-433) x US-10-314-657-1 (1-135638)
Qy 6 GylileVallysAlaAspValAlaGlnAsnAsn----- 16
Db 8067 GGCAGGGTGAAGCCGATCTGGCCGACTCCACGCCAGATCGGCGCGGAGGATATGG 8126
Qy 17 ---TyGlyLeuTyrGlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThr 35
Db 8127 GCGAGGGCCACACCGCCGACGAGTGAAGTCCGATCGTCCACAGCGCGGACAC 8186
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysile-----ThrAlaLeu 53
Db 8187 -----GNACACCCGACCTGTCGGCAGGTGTCGACGCGGCGCAGC 8228
Qy 54 TyrAlaLeuGlyArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAla 73
Db 8229 TTGCTCCCGCGGAGGACGACATCGCC--GACTACACGCGGCGGACGCGCGCGCC 8285
Qy 74 GlySerValLeuGly-----AsnAlaLeuAsnLysGlyMetAlaProGln 88
Db 8286 TCAGACCATCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8345
Qy 89 AlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSer 108
Db 8346 GCGCGGCTGCTCGCGCAAGGTGCTCACTCCGAGGCG-----AGCGCGCGAGGAATCG 8399
Qy 109 AsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAla-----GlyAlaArgIleHis 125
Db 8400 TGGATC-----ATCGCGGCGCATGGAGTGGCGCGCGCGCGCGGCGGCGGCGGCGG 8453
Qy 126 ThrAsnSerTrpGlyAlaProValAsnGlyValaTyrThrAlaAsn-----SerArg 142
Db 8454 AGCATGAGCCTGGGC-----GGCGGCGGTGACAAAGAACGACCGCGATGAGCGAG 8501
Qy 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 8502 GCGGTCCAGCAACTCAGCCACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8561
Qy 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 8562 GGCGCGGCGACTCC-----ATCAGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8615
Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla 202
Db 8616 GCC-----GTGCGACTCCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8642
Qy 203 GlnPheSerSerArgGly-----AlaThrArgAspGlyArgIleLysProAspValThrAla 221
Db 8643 GACTTCTCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8702
Qy 222 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAla 241
Db 8703 CCGCGGCGTGCACATCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8756
Qy 242 AsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAla 261
Db 8757 -----TACACCGCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8801
```

RESULT 5

```

US-10-314-657-1
; Sequence 1, Application US/10314657
; Publication No. US20030175888A1
; GENERAL INFORMATION:
; APPLICANT: SHEN, Ben
; APPLICANT: CHENG, Yi-Qiang
; APPLICANT: TANG, Gong-Li
; TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
; FILE OF INVENTION: Syntheses and Methods of Use
; FILE REFERENCE: 054030-0021
; CURRENT APPLICATION NUMBER: US/10/314,657
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: PCT/US02/08937
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/278,935
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 214
```

QY 262 GlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyLeuThrProLysPro 281
Db 8802 GGTGTCCGCGCTCTCGCCGCGAGCAGCCCGAGTGGACCGGACCCAGCTCAAGGAG 8861
QY 282 SerLeuIleLysAlaLeuLeuAlaGlyAlaThrAspValGlyLeuGlyTyrProSer 301
Db 8862 GCAGTGGTCAAGCAGCCCAAGCAACCGCGCG- -TACACCCCG 8903
QY 302 GlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaLysValAsn 321
Db 8904 TACCAGCGGCGCGCGCTCGACGCGCGCGCGCGCTGCACACCGAGTCTTCGCC 8963
QY 322 GluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAlaGly 341
Db 8964 ACCACGACCGCTTACTCC- -GGCTTCCAC- - 8990
QY 342 LysProLeuLysIleSerLeuValThrAspAlaProGlySerThr- -Thr 358
Db 8991 - - - - -ACGTGCCCCCAAGCCCGGGGAGACCGATCTCCGGAG 9029
QY 359 AlaSerTyrThrLeuValAsnAsp- - - - -LeuAspLeuValle- -Thr 372
Db 9030 GTACGCTACACCAACGTGCGGACGCGCGCTCAACCTGCGCTCAACGCGCAC 9089
QY 373 AlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyr 392
Db 9090 GTCCGCGCGGGTG- - - - -TTCAGCCTCTCCGAGACCAT- - 9125
QY 393 AspGlyArgAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThr 412
Db 9126 - - - - -GTCACCGTGCCTGCGCGGACCGCACCGCGCAC 9155
QY 413 IleGluVal- - - - -GlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
Db 9156 GTACCCCTGACCGCGCTCTGGCAAGAGTGGCGGGGACCGATCGTGGTCAAGCGCGGTGATC 9215

RESULT 6

US-09-974-300-1934
; Sequence 1934, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berkka, Randy M.
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085,500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1934
; LENGTH: 1329
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1934

Alignment Scores:

Pred. No.:	1,08e-26	Length:	1329
Score:	334.00	Matches:	105
Percent Similarity:	44.58%	Conservative:	43
Best Local Similarity:	31.63%	Mismatches:	116
Query Match:	14.77%	Indels:	68
DB:	9	Gaps:	12

US-09-985-689A-5 (1-433) x US-09-974-300-1934 (1-1329)

QY 11 AspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAsp 30
Db 406 GAAGTGGTCAAGAAACATCAGACGCTGACAGGCAAGGAGTGACAGTGGTGTGATGAT 465

RESULT 7

US-10-156-761-3306
; Sequence 3306, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

```

; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIORITY FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3306
; LENGTH: 3417
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3417)
US-10-156-761-3306

Alignment Scores:
Pred. No.: 15e-25 Length: 3417
Score: 329.00 Matches: 114
Percent Similarity: 45.2% Conservative: 44
Best Local Similarity: 32.66% Mismatches: 126
Query Match: 14.55% Indels: 67
DB: 14 Gaps: 12

US-09-985-689A-5 (1-433) x US-10-156-761-3306 (1-3417)

QY 18 GlyLeuTyrGlyGlnGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
DB 745 GGGTACGACGGAAGGGCGCTCAAGATCGCGCTCTGGACACGGGTGTCGAC----- 795
QY 38 AsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
DB 796 -----GCGACCCACCGGACCTCAGGACCGAGTGGCGGAGTCCAGAGACTTCTCC 846
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
DB 847 GCGCGCGCGGCGCGCGCGGACCATCTCGGTCAAGGCGGCGGCGGCGGCGGCGGCGG 906
QY 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
DB 907 GGCACCGCGCGCAAGTCCACGCGCAAGTACAAAGGTGTGCGCGCGGCGGCGGCGGCGG 966
QY 93 PheGlnSerIleMetAspSerSerGly-----GlyLeu 103
DB 967 AACGGCAAGTCTCTCAGCACACACCGGCTCCGGCGGACGACTCCGGCATCTCGCGGCGATG 1026
QY 104 -GlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaThrAsnAlaGlyAlaAr 123
DB 1027 GAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1086
QY 123 gilleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArgG1 143
DB 1087 CCGGAGATCGACCCGCTGGAA-GCGGAGGTCAACAGCTC----- 1125
QY 143 nValAspGluTyrValArgAsnAsnAspMetThrValLeuPhe-----AlaAlaGlyAs 161
DB 1126 -----TCCAGAGAGAGGCGATCTCTTCGGCATCGCGCGCGGCGGCA 1166
QY 161 nGluGlyPro---AsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleTh 180
DB 1167 CGAAGCGGATTCGGCGGAGACAGACCATCGGCTCCCGGCGGCGGCGGCGGCGGCGGCG 1226
QY 180 rValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHi 200
DB 1227 CGTCGCGCGC-----GTGAACGACGACGACAA 1253
QY 200 sileAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValTh 220
```

```

DB 1254 GCTGGCGTCTCTTCCAGCGCGGCGCGCGCTGGACGCGCGCATCAAGCGCGGCTCAC 1313
QY 220 xAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTr 240
DB 1314 CGACCCGCGGTGGACATCACCGCGC-----GCGGCGCGGCGGCGGCGGCGGCGG 1364
QY 240 pAlaAsnTyrAsnSerLys-----TyrAlaTyrMetGlyGlyThrSerMetAlaTh 257
DB 1365 CCAGGAGGTGCGGCGGAGCGGAGCGGTACCTCACCATCTCCGTAGCTCGATCGCGGAC 1424
QY 257 rProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyI1 277
DB 1425 CCGCATGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1467
QY 277 eThrProLysProSerLeu-----IleLysAlaAlaLeuIleAlaGlyAlaThrAspVa 295
DB 1468 -----CCAACTGCTGCTCGCGAGCTCAAGGCGCGCTGACCGGCTCGCGGAAG----- 1518
QY 295 lGlyLeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLe 315
DB 1519 -GCGGCAAGTACACCGCGTTCAGCGGCGTCCGCGGCGGCGGCGGCGGCGGCGGCGGAT 1577
QY 315 uAsnValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaAlaThrTyrSe 335
DB 1578 CAAGCAGTCCGTGATCGCCACCGCAACTCGGTGAGCTTCGGCAT-CCAGCAGTGGCGGC 1636
QY 335 rPheGlnAlaGlnAlaGlyLysPro 343
DB 1637 ACACCGACGACAAAGCGGTCACCCA 1661

RESULT 8
US-10-156-761-1/C
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HORIKAWA, JUN
; APPLICANT: HIRAKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIORITY FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 1.37e-20 Length: 9025608
Score: 329.00 Matches: 114
Percent Similarity: 45.2% Conservative: 44
Best Local Similarity: 32.66% Mismatches: 126
Query Match: 14.55% Indels: 67
DB: 14 Gaps: 12

US-09-985-689A-5 (1-433) x US-10-156-761-1 (1-9025608)

QY 18 GlyLeuTyrGlyGlnGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
```

Db 4132505 GGGTACAGCGCAGGGCGTCAAGATCGCGTCTCGACACCGGTCTGCAC----- 4132455
Qy 38 AsnAspSerMetHisGluAlaPheArgGlyLeuThrAlaLeuValLeuGly 57
Db 4132454 -----CGACACCGCGGACCTCAAGACCGAGTGGCCGAGTCCAGAACTTCTCC 4132404
Qy 58 ArgThrAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 4132403 GCG 4132344
Qy 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 4132343 GGCACCG 4132284
Qy 93 PheGlnSerIleMetAspSerGly-----GlyLeu 103
Db 4132283 AACGGCAAGGTCTCGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4132224
Qy 104 -GlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaAr 123
Db 4132223 GAGTGGCG 4132164
Qy 123 GileHisThrAsnSerTrpGlyAlaProValAsnGlyAlaThrAlaAsnSerArgG 143
Db 4132163 CCGGAGATCGACCG 4132125
Qy 143 nValAspGluTrpValArgAsnAsnAspMetThrValLeuPhe-----AlaAlaGlyAs 161
Db 4132124 -----TCCGAGGAGAGGGCGCATCTCTTCGCGATCGCGCGCGCGCGCGCGCG 4132084
Qy 161 nGlyGlyPro---AsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaLeu 180
Db 4132083 CGAAGGCGAGTTCGCGGAGCGAGACCATCGGTCTCCCGCGGAGCGCGCGCGCGCGCG 4132024
Qy 180 rValGlyAlaThrGluAsnTrpArgProSerPheGlySerLeuAlaAspAsnProAsnHi 200
Db 4132023 CGTGGCGCGCG-----GTGAACGACGACGCGCGCGCGCGCGCGCGCGCGCG 4131997
Qy 200 sIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 220
Db 4131996 GCTGGCGTCTCTTCCAGCG 4131937
Qy 220 rAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPheTr 240
Db 4131936 CGCACCGCGGTGGACATCATCGCGCGCG-----CGCGCGCGCGCGCGCGCGCGCGCG 4131886
Qy 240 pAlaAsnTrpAsnSerLys-----TyrAlaTrpMetGlyGlyThrSerMetAla 257
Db 4131885 CCAGGAGGTGGCG 4131826
Qy 257 rProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGly 277
Db 4131825 CCGGATGTCG 4131783
Qy 277 eThrProLysProSerLeu-----IleLysAlaAlaLeuIleAlaGlyAlaThrAsp 295
Db 4131782 -----CCCAACTGTCGTCGCGAGCTCAAGGGCGCGGTGACCGCGCGCGCGCGCG 4131732
Qy 295 lGlyLeuGlyTrpProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSer 315
Db 4131731 -GGCGGCGAGTACACCGCGTCCAGCGAGGTTCGGCGCGGTATCGCGCGCGCGCGCGCG 4131673
Qy 315 uAsnValAlaTrpValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTrpSe 335
Db 4131672 CAAGCAGTCCGCGTATCG 4131614
Qy 335 rPheGlnAlaGlnAlaGlyLysPro 343
Db 4131613 ACACCGACGACGCGCGGTTCACCGCA----- 4131589

RESULT 9

US-10-156-761-5701

; Sequence 5701, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HOSIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5701
; LENGTH: 3624
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3624)
US-10-156-761-5701

Alignment Scores:
Pred. No.: 2,71e-25 Length: 3624
Score: 327.00 Matches: 119
Percent Similarity: 42.75% Conservative: 49
Best Local Similarity: 20.28% Mismatches: 161
Query Match: 14.46% Indels: 64
DB: 14 Gaps: 16

US-09-985-689A-5 (1-433) x US-10-156-761-5701 (1-3624)

Qy 8 ValLysAlaAspValAlaGlnAsnAsn-----Tyr 17
Db 556 GTGAGGCGCGCATGCGCGAGGACCGCGAGATCGGTACGCGCGCGCGCGCGCGCGCGCG 615
Qy 18 GlyLeuTrpGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 616 GGGCTCACG 669
Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 670 -----ACTACCG 717
Qy 58 ArgThrAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 718 GACGGGAGGAGGTGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 777
Qy 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 778 GGCAGCG 837
Qy 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThr 112
Db 838 GTGCGCAAGTGTCTCAGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 885
Qy 113 LeuPheSerGlnAlaTrpAsnAla-----GlyAlaAsnGlyHisThrAsnSerTrp 129
Db 886 ATCGCGCGCATGAATGGCG 945
Qy 130 GlyAla---ProValAsnGlyAlaTrpThrAlaAsnSerArgGlnValAspGluTrpVal 148
Db 946 GGATCGACCG 1005
Qy 149 ArgAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThr 168
Db 1006 GAGGAGACCG 1062

QY 169 IleserAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArg 188
Db 1063 ATCGCTCGCCGGCGCCGACCTCCGCGTGCAGCGCGCC----- 1107
QY 189 ProSerPheGlySerLeuAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGly 208
Db 1108 -----GTCGACTCATCCAGCCGCGCTACTTTCACCGCGCGCC 1149
QY 209 AlaThrArg---AspGlyArgIleLysProAspValThrAlaProGlyThrPheIleLeu 227
Db 1150 CGCGCCCGCGGCGACACCGCTCAAGCCCGGCTCGCGCGCGCGCGTGCATCCGC 1209
QY 228 SerAlaArgSerSerLeuAlaProAspSerPheTyrAlaAsnTyrAsnSerLysTyr 247
Db 1210 GCGGCGCGCTCCAGCTCGCCCGCGCGCGCTAC-----TAC 1248
QY 248 AlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValaGlnLeu 267
Db 1249 ACCTCCATGAGCGGTACGTGATGCGAGCGCCCATGTGCGGGGGTGCAGCGCTCCTC 1308
QY 268 ArgGluHisPheIleLysAsnArgGlyIleThrProLysProSerLeuIleLysAlaAla 287
Db 1309 GCGGAGCGACCGCGACTGAGCGCGCGCGCTCAGAGCGCGCTGATGTCACGCTCC 1368
QY 288 -----LeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrProSerGlyAspGlnGly 305
Db 1369 GAGCAACTCGAGCGCTCCGTATATCAGTTGGGGGGGTGCGGTGAGTGTGCGGAGCGCC 1428
QY 306 TrpGly---ArgValThrLeuAspLysSerLeuAsnValAlaTyr-----ValAsn 321
Db 1429 GTCGCGCGCGCGTCCAGCGCGCGCGCGCGCGCGCGCTCGGCTCCACCGTGGCGCCAT 1488
QY 322 GluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAlaGly 341
Db 1489 GAGCGCGACTCGCCCGTCAG-----AAGACGCTCACCTACTCCAGCTCCCTCCGAC 1539
QY 342 LysProLeuLysIleSerLeuValThrPheAspAlaProGlySerThrAlaSerTyr 361
Db 1540 ACGAGCTCGAGTTGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1590
QY 362 ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 374
Db 1591 ACCTCGCC-----GACACCGCACTCACCGTGGCC 1620

RESULT 10
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)

; OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1
Alignment Scores: 2.28e-20 Length: 9025608
Pred. No.: 327.00 Matches: 119
Score: 42.75% Conservative: 49
Best Local Similarity: 30.28% Mismatches: 161
Query Match: 14.46% Indels: 64
DB: 16 Gaps: 16
US-09-985-689A-5 (1-433) x US-10-156-761-1 (1-9025608)
QY 8 ValIleAlaAspValAlaGlnAsnAsn-----Tyr 17
Db 6918813 GTCGAGCGGACATGCGCGGAGAGCAACGCGCAGATCGTACGCGGCGCGGTGGGAGCGCC 6918872
QY 18 GlyLeuTyrGlnGlnGlnValValAlaAlaAspThrGlyLeuAspThrGlyArg 37
Db 6918873 GGGCTCAGCGGCGACCGCGCTACCGTCCGCGTCTCGACACCGCGGTGCACACC----- 6918926
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 6918927 -----ACTCACCCTCGCGCGCGCGGTCTCCCGGAGCAAGAGCTTCATC 6918974
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 6918975 GACGGGAGGAGGTGCGCGACCGCAACGCGCACCGGACCGCTCACCTCGACCGTCCGC 6919034
QY 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 6919035 GGCAGCGCGCGCGCTCCGACGCGACGAGCGCGGTGCGCGCGCGTCCGCGTCCGCTCGCC 6919094
QY 93 PheGlnSerIleMetAspSerSerGlyLeuGlyLeuProSerAsnLeuAsnThr 112
Db 6919095 GTCGGCAAGTGTCTCAGCGACCGAGCGCG-----GGAAGCGAGTCCAGATC 6919142
QY 113 LeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHisThrAsnSerTyr 129
Db 6919143 ATCGCGGCGATGAAATGGCGCGCGCGCGCGCGTCCGCGGAGTCTCTCGAGAGCTTC 6919202
QY 130 GlyAla---ProValAsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGlyTyrVal 148
Db 6919203 GGATCGACCGGAGCGCGGAGCGGACCGACCGCGCGCGCGCGCGTCCGACACCTCTCC 6919262
QY 149 ArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyTyr 168
Db 6919263 GAGGAGACCGCGCGCTCTTCGTCGTCGCGCGGGAACACACCGGTCCCGCTCC-----TCG 6919319
QY 169 IleserAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArg 188
Db 6919320 ATCGCTCGCGCGCGCGCGCTCTTCGTCGTCGCGCGGGAACACACCGGTCCCGCTCC 6919364
QY 189 ProSerPheGlySerLeuAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGly 208
Db 6919365 -----GTCGACTCATCCGCGCGCGCGCTTCTCCGCGCGCGCGCGCGCGCGCGCG 6919406
QY 209 AlaThrArg---AspGlyArgIleLysProAspValThrAlaProGlyThrPheIleLeu 227
Db 6919407 CGCGCGCGCGGACACCGCGCTTCAGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCG 6919466
QY 228 SerAlaArgSerSerLeuAlaProAspSerPheTyrAlaAsnTyrAsnSerLysTyr 247
Db 6919467 GCGCGCGCTCCAGCTCGCGCGCGCGCGCTAC-----TAC 6919505
QY 248 AlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 267
Db 6919506 ACCTCCATGAGCGGTACGTGATGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919565
QY 268 ArgGluHisPheIleLysAsnArgGlyIleThrProLysProSerLeuIleLysAlaAla 287
Db 6919566 GCGGAGCGACCGCGCTGAGCG 6919625

QY 288 -----LeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrProSerGlyAspGlnGly 305
Db 6919626 GAGCAACTCAGCGCTCCGATATCATAGTTCGGGGCGGTCGGTCAAGTTCGGGAGCGC 6919685
QY 306 TrpGly---ArgValThrLeuAspLysSerLeuAsnValAlaTyr-----ValAsn 321
Db 6919686 GTGGGCGCGCGTCAACGGAGCGGCGAGCGCGCTCCGCTCCACCGCTGGCCCAT 6919745
QY 322 GluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAlaGly 341
Db 6919746 GAGCCGATCGACCGTCACG-----AAGCGGTACCTTACTTCCAACTCCCTCCGAC 6919796
QY 342 LysProLeuLysLysSerLeuValTrpThrAspAlaProGlySerThrThrAlaSerTyr 361
Db 6919797 ACGACGTCGAGTTCAGCTCGCGTTCGGGGCGCGCGCGCGTTCGCGC----- 6919847
QY 362 ThrLeuValAsnAspLeuAlaValLeuAlaPro 374
Db 6919848 ACCCTCGCC-----GACACCGCACTCACCGTGCCC 6919877

RESULT 11

US-10-090-624-5
; Sequence 5, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 4765
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-5

Alignment Scores:
Pred. No.: 6,41e-23 Length: 4765
Score: 307.00 Matches: 129
Percent Similarity: 36.48% Conservative: 37
Best Local Similarity: 28.35% Mismatches: 153
Query Match: 13.58% Indels: 136
DB: 13 Gaps: 15

US-09-985-689A-5 (1-433) x US-10-090-624-5 (1-4765)

QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyA:GAsnAspSer 40
Db 1024 GCAATGTTATGACATTGCATATGTGTGATACCTTGCATCTGACTGACCTTCACCGACGAA 1083
QY 41 -----SerMetHisGluAlaPheArgGly 48
Db 1084 GTTCCACTTGGCCAGTACACGTTACTTATGATGTTGCTGTTTTTAGCTACTACTACGGT 1143
QY 49 LysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn----- 66
Db 1144 CCTCTCAAC-----TAGTGTCTCCAGAAATA-----GATCCTTACGGAGAA 1185
QY 67 -----GlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1186 TATGCAATATTGGGTGGGATGGTCAACGGTCAACGAACTCAGTAGCTGGAACCTGTTGCT 1245

QY 78 GlyAsnAlaLeuAsn----- 82
Db 1246 CGTTACGACGACCAACAATGATGCTTGGGATTGGCTCAGTATGTAATCTGCTGTAATGGAA 1305
QY 83 -----LysGly 84
Db 1306 GTGTTCTCAGACCTCTATGTTGGGATTATACGAACGTTTACCACAGACACCGTSCAGGGT 1365
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGly 104
Db 1366 GTGCTCCAGGTGCCAAATATGCAATAGAGTTCCTTAGGATGTATGA-----CGG 1419
QY 105 GlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIle 124
Db 1420 GGTAGCATGTGGGATATTATAGAAGGTATGACATACGACCAACCCATGTTGTCAGACGTT 1479
QY 125 HisThrAsnSerTrpGly-----AlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
Db 1480 ATAAGCATGAGTCTCGGTGGAAATGCTCCATCTAGTGTACTGATCCAGAAAGCGTT 1539
QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 1540 GCTGTGGATGAGTTCACGAAAGTACGGTGTGTTATTCGTAATAGTCTGCAGGAAATGAA 1599
QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 1600 GGTCTCGCATTAACATCGTTGGAGTCTCTGTTGTCACAAAGGCAATACCTGTGGA 1659
QY 183 AlaThrGlu----- 185
Db 1660 GCTGCTGCAGTGCCTCCATTACGTTGGAGTTTATGTTTCCCAAGCACTGGGATATCTCAT 1719
QY 186 -----AsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIle 201
Db 1720 TACTATGATTTCTATTCTCCCGCTTACACAAACGTT-----AGAATA 1764
QY 202 AlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAla 221
Db 1765 GCATTCTTCTCAAGCAGAGGCGCGAGATAGATGTTGAAATAAAACCAATGATGTGCT 1824
QY 222 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAla 241
Db 1825 CCAGGTTACGGAATTTACTCATCTCCCGGATGGATGGCGGAGCTGACTTC----- 1878
QY 242 AsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAla 261
Db 1879 -----ATGTCGGAATTCGATGGCTACTCCACATGTTCAGC 1914
QY 262 GlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysPro 281
Db 1915 GGTGTGTTGTCACCTCTCATAGCGGG---GCCAAGGCGCGGGAATATATCAATCCA 1971
QY 282 SerLeuIleLysAlaAlaLeuIleAlaGlyAlaThr-----AspVal 295
Db 1972 GATATAATTAGAAGTTCCTTGAGAGCGGTGCACCTGGCTTGAGGAGATCCATATACT 2031
QY 296 GlyLeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeu 315
Db 2032 GGGCAGAAGTACACTGAGCTTGACCAAGTCTGTTGTTTAACTTACCAAGTCCCTGG 2091
QY 316 AsnValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSer 335
Db 2092 GAAATC----- 2097
QY 336 PheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGly 355
Db 2098 CTTAAGGCTATAAAGCGCACCACTCTCCCAATTGTTGATCACTGGCGCAGACCAAGTCTTAC 2157
QY 356 SerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsn 375
Db 2158 AGCGAATTTGGCGAGTACTGGGTGTG---GACGTTATAGAGGTCTCTACGCCAGGAAAC 2214


```
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 113
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: avix16 dna: partial sequence of coding strand 1; nucleotide 1
; OTHER INFORMATION: corresponds to nucleotide 56,198 of coding strand 1.
US-10-084-846A-113

Alignment Scores:
Pred. No.: 2,66e-22 Length: 1560
Score: 295.00 Matches: 125
Percent Similarity: 42.11% Conservative: 51
Best Local Similarity: 29.90% Mismatches: 153
Query Match: 13.05% Indels: 90
DB: 15 Gaps: 19

US-09-985-689A-5 (1-433) x US-10-084-846A-113 (1-1560)

Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 451 GGGCAGGAGTACCGCGGTACGTCATCGACACCGCGCTC-----CGC 492

Qy 41 SerMetHisGluAlaPheArgGlyLysLeuThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 493 ATCACCCACAGCGATTCGGCGCGCGCGCTCTCTACCGCTACGACGCCATCGCAACAGCAG 552

Qy 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 553 AACACGCCCGAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 612

Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerLeuMetAspSer 99
Db 613 GCCTTAC-----GGCGTCGCGCAAGAGCAAGATCGTAGCGCTCGCGTGTGTAACAC 666

Qy 100 Ser-----GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
Db 667 TCCGGCAGGGCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 726

Qy 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro-ValAsnGlyAlaTy 137
Db 727 GT-CAAGCGCGCGCTCGCCACAT-----GTCCCTCGCGCGCGCGCGCGCGCGCG 770

Qy 137 rThrAlaAsnSerArgGlnValAlaAspGluTyrValArgAsn-----AsnAspMetTh 154
Db 771 CACGGCC-----CTCGACACGGCGGTACGCAACGCCATGCGCTCCCGCGGTAC 818

Qy 154 rValLeuPheAlaAlaGlyAsnGlyGlyProAsnSerGlyThrIleSerAlaProGlyTh 174
Db 819 CTTCGCGGTGGCGCGCGACAGTACGACCAACGCTCCACGAGGTCA---CCCGCAGC 875

Qy 174 rAlaLysAsnAlaLeuThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLe 194
Db 876 CGTCACCGAGGCCATACCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 930

Qy 194 uAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyAr 214
Db 931 -----TACTCCAACTACGGCTCCGCTCCTC----- 954

Qy 214 GileLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAl 234
Db 955 -----GACCTCTTCGCGCGCGCGGTCTGCTTCCATCACTTCGGCC----- 990

Qy 234 aproAspSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSe 254
Db 991 -----TGAACCTACGCGACTCGCGGACCAACACCATCTCCGCTACGTC 1034

254 rMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAs 274
1035 GATGGCAGCCCGCACGTGGCGCGCGCGCGCGCTC-----CACCTCGCGCGCAA 1085
274 nArgGlyIleThrProLys-----ProSerLeuIleLysAlaAlaLeuIleAlaG 291
1086 CCCCTCGGCCACCCCGCTCCAGGTGCGCACGGCGCTGACGTCCGCGCGCCACCGCGCT 1145
291 yAlaThrAspValGlyLeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLe 311
1146 CGTCACCAACCCCGCGCGCGCTCGCCCAAC----- 1176
311 uAspLysSerLeuAsnValAlaValAsnGluAlaThrAlaLeuAlaThrGlyGlnLy 331
1177 -----CGGCTCCTGTACGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1223
331 s-----AlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLy 345
1224 CTTGAGAACACCGGTGACTACGATCAGCAGCACTCCACGCTCGAGTCCCGGTGAC 1283
345 sIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----Ty 361
1284 GGTCTCCGGCGTCTCCGGCAACGGCGCTCGGCCCTCGCGCTAGAGGTCCACATCGCTCA 1343
361 rThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGl 381
1344 CACGTACATCGCGACCTCCAGGTCCAGTCCAGTCCCGCGCGCGCGCGCGCGCTACCGCT 1403
381 yAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnValGluAsnVa 401
1404 CAAG-----TCGTACGCGCACCGCGCGGTTCGGAACAATCAACACCGAGTACTCGGT 1457
401 lPheIleAsnAlaProGln-----SerGlyThrTyrThrIleGluVal 415
1458 G-----AACGCTCTCTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTG 1503

RESULT 14
US-10-084-846A-1
; Sequence 1, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: TRETZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: PCT/EP01/09815
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: US/10/084,846A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 1
; LENGTH: 59816
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-1

Alignment Scores:
Pred. No.: 5.25e-20 Length: 59816
Score: 295.00 Matches: 125
Percent Similarity: 42.11% Conservative: 51
Best Local Similarity: 29.90% Mismatches: 153
Query Match: 13.05% Indels: 90
DB: 15 Gaps: 19

US-09-985-689A-5 (1-433) x US-10-084-846A-1 (1-59816)

Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
```

Db 56648 GGGCAGGAGTACGGCGTACCTCATCGACACGGCGTC-----CGC 56689
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 56690 ATCACCCACAGGACTTCGGCGCGCGGCTCTCTACGGGTACGACGCATCGACACAGC 56749
Qy 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 56750 AACACGCCAGCAGCGCCGACGCGCCACGCGACGCGCTGCGCGGCGCGGCGTCCGCGCAAC 56809
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 56810 GCCTAC-----GGCGTCGCCAAGAGGCGCAAGATCGTAGGGTCCGGGTGTGAACAAC 56863
Qy 100 Ser-----GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
Db 56864 TCCGGCCAGCGCACACCCCGCCAGGTCTGCGCGGCATCGACTGGGTGCGCCGCGAACGCC 56923
Qy 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro-ValAsnGlyAlaTy 137
Db 56924 GT-CAAGCGCGCGCTCGCCACAT-----GTCCCTCGCGCGCGCGCGCA 56967
Qy 137 rThrAlaAsnSerArgGlnValAspGluTyrValArgAsn-----AsnAspMetTh 154
Db 56968 CACGGCC-----CTCGACACGGCGGTACGACCAACGCGATGCGCTCGGGGTGCAC 57015
Qy 154 rValLeuPheAlaAlaGlyAsnGlyProAsnSerGlyThrIleSerAlaProGlyTh 174
Db 57016 CTTGCGCGTGGCGCGCGCGAACAGTCGACCAACGCTCCACGAGTCA---CCGCGCAG 57072
Qy 174 rAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIe 194
Db 57073 CGTCACCGAGGCATCACGGTCGGCGCGACGACGCTCGGACGCCAAGCGCGGC----- 57127
Qy 194 uAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyAr 214
Db 57128 -----TACTCCAACTACGGCTCCGTCCTC----- 57151
Qy 214 gIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAl 234
Db 57152 -----GACTCTCTGCCCCCGGTTCTGTCATCATCCTCGGC----- 57187
Qy 234 aproAspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSe 254
Db 57188 -----TGGAACTCAAGCGACTCGGCGGACCAACACCATCTCCGGTACGTC 57231
Qy 254 rMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAs 274
Db 57232 GATGGCAGCCCGCACGTGGCGGCGCGCGCTC-----CACCTCGCGCCAA 57282
Qy 274 nArgGlyIleThrProLys-----ProSerLeuIleLysAlaAlaLeuIleAlaGl 291
Db 57283 CCCCCTCGGCCACCCCGTCCAGGTGCGCACGCGGCTGACGTCCGCGCCACACCGCGGT 57342
Qy 291 yAlaThrAspValGlyLeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLe 311
Db 57343 CGTCACCAACCCCGGACGGGCTCGCCAAC----- 57373
Qy 311 uAspLysSerLeuAsnValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLy 331
Db 57374 -----CGGCTCTCTGTACGTGCGCGCGCGCGCACACCCCTCCGCGCGCGCGG 57420
Qy 331 s-----AlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLy 345
Db 57421 CTTGCGAACAACCGGTGACTACATCAGGACCACTCCACGGTGAAGTCCCGCGTGCAC 57480
Qy 345 sIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----Ty 361
Db 57481 GGTCTCGGGGTCTCCGCGCAACGGCGCTCGGCGCTCGCGCTAGAGTCCACATCGTCCA 57540
Qy 361 rThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGl 381
Db 57541 CACGTATACGGCACCTCCAGGTCACGCTGATCGCCCCCGACGCGCACGCGTACAGCT 57600

Qy 381 yAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnVa 401
Db 57601 CAAG-----TCGTACGGCACCGCGCGCATTCGGACAACATCAACACACGACTACGGT 57654
Qy 401 lPheIleAsnAlaProGln-----SerGlyThrTyrThrIleGluVal 415
Db 57655 G-----AACGCCTCTCGGAGCGCGCACGCGACGTCGGAACCTCGGGGTG 57700
RESULT 15
US-10-084-846A-2/c
; Sequence 2, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WELTNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; PCT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent In Ver. 3.2
; SEQ ID NO 2
; LENGTH: 59816
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-2

Alignment Scores:
Pred. No.: 5,25e-20 Length: 59816
Score: 295.00 Matches: 125
Percent Similarity: 42.11% Conservations: 51
Best Local Similarity: 29.90% Mismatches: 153
Query Match: 13.05% Indels: 90
DB: 15 Gaps: 19

US-09-985-689A-5 (1-433) x US-10-084-846A-2 (1-59816)

Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 3169 GGGCAGGAGTGACGGCGTACGTATCGACACCGCGCTC-----CGC 3128
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 3127 ATCACCCACAGGACTTCGGCGCGCGCTCTCTACGGTACGACGCATCGACACAGC 3068
Qy 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 3067 AACACGCCAGGACGCGCCACGCGCACGCGCACGCTGCGCGCGCGCGTCCGCGCAAC 3008
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 3007 GCCTAC-----GGCGTCGCCAAGAGGCGCAAGATCGTAGGGTCCGGTGTGAACAAC 2954
Qy 100 Ser-----GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
Db 2953 TCCGCCACAGGCGCACCCACCGCGCGCATCGTCCGCGCATCGACTGGGTGCGCCGCGCAACGCC 2894
Qy 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro-ValAsnGlyAlaTy 137
Db 2893 GT-CAAGCGCGCGCTCGCCACAT-----GTCCCTCGCGCGCGCGCGCA 2850
Qy 137 rThrAlaAsnSerArgGlnValAspGluTyrValArgAsn-----AsnAspMetTh 154
Db 2849 CACGGCC-----CTCGACACGGCGGTACGCAACGCGCATGGCTCCGGCGTGCAC 2802
Qy 154 rValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyTh 174

Search completed: April 5, 2004, 00:57:57
Job time : 6868.25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:40:05 ; Search time 2274.58 Seconds
(without alignments)
5684.703 Million cell updates/sec

Title: US-09-985-689A-5
Perfect score: 2261
Sequence: 1 NDVARGIVKADVAQNNGLY.....EVQAYNVSPQRFSLAIVH 433

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-O/cgn2_1/USPTO.spool/US09985689/runat_31032004_161807_4168/app_query.fasta_1.3498
-DS=EST -OFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 @CNG_1_1_7084 @runat_31032004_161807_4168 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	235.5	10.4	1605	13	BQ622771	BQ622771 CC Contig
2	228	10.1	4198	11	AK029048	AK029048 Mus_muscul
3	215.5	9.5	594	12	BQ393752	BQ393752 BQ393752
4	214.5	9.5	640	12	BQ395336	BQ395336 BQ395336
5	204	9.0	2141	13	BQ142519	BQ142519 Contig6 M
6	199.5	8.8	532	29	TA319G10P	AL492464 T. brucei
7	199.5	8.8	718	12	BI750157	BI750157 Fg02_10G0
8	195.5	8.6	771	14	CA320325	CA320325 UI-M-FW0-
9	193.5	8.6	601	12	BQ387574	BQ387574 BQ387574
10	190	8.4	545	13	BQ575479	BQ575479 T9ESTZyb8
11	188	8.3	574	29	TA315H10P	AL490202 T. brucei
12	188	8.3	614	9	AJ273402	AJ273402 AJ273402
13	183.5	8.1	2121	28	BZ424955	BZ424955 100023066
14	182.5	8.1	665	13	BQ770462	BQ770462 UI-M-FIO-
15	182.5	8.1	716	28	BZ893395	BZ893395 HL2_0177
16	180	8.0	530	29	CNSC10FO	AL153820 Anopheles
17	180	8.0	650	9	AJ274038	AJ274038 AJ274038
18	179.5	7.9	675	14	CF727824	CF727824 UI-M-HB0-
19	178	7.9	508	28	AQ652212	AQ652212 Sheared D
20	178	7.9	633	12	BQ369190	BQ369190 BQ369190
21	177.5	7.9	576	14	CD295943	CD295943 StrPu691.
22	176.5	7.9	1002	29	CNS0606B	AL393417 T3 end of
23	176.5	7.8	594	14	CF846826	CF846826 psHB0362A
24	176.5	7.8	1029	29	CNS071DW	AL424794 T7 end of
25	175	7.7	580	9	AJ273745	AJ273745 AJ273745
26	175	7.7	583	9	AJ273947	AJ273947 AJ273947
27	175	7.7	593	9	AJ273918	AJ273918 AJ273918
28	175	7.7	601	9	AJ273050	AJ273050 AJ273050
29	175	7.7	601	9	AJ273921	AJ273921 AJ273921
30	173.5	7.7	641	12	BQ393925	BQ393925 BQ393925
31	172	7.6	610	9	AJ274218	AJ274218 AJ274218
32	170	7.5	564	28	AQ651427	AQ651427 Sheared D
33	169.5	7.5	937	12	BQ246418	BQ246418 602360428
34	169	7.5	609	9	AJ272712	AJ272712 AJ272712
35	169	7.5	794	14	CF737198	CF737198 UI-M-HD0-
36	168	7.4	781	13	BQ612128	BQ612128 UI-M-EW0-
37	167.5	7.4	604	9	AJ274126	AJ274126 AJ274126
38	167.5	7.4	615	9	AJ274059	AJ274059 AJ274059
39	167	7.4	616	13	CF138007	CF138007 UI-HF-BNO
40	166.5	7.4	895	13	BQ216158	BQ216158 AGENCOURT
41	166	7.3	619	9	AJ273097	AJ273097 AJ273097
42	166	7.3	681	14	CB690041	CB690041 CBST-54-B
43	166	7.3	887	13	BQ879057	BQ879057 AGENCOURT
44	165.5	7.3	593	9	AJ273903	AJ273903 AJ273903
45	165.5	7.3	596	9	AJ273379	AJ273379 AJ273379

ALIGNMENTS

RESULT 1
BQ622771
LOCUS
DEFINITION
BQ622771
CC Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
CDNA, mRNA sequence.
1605 bp mRNA linear EST 01-JUL-2002
ACCESSION
BQ622771.1 GI:21649940
VERSION
EST.
KEYWORDS
Conidiobolus coronatus (Delacroixia coronata)
SOURCE
Conidiobolus coronatus
ORGANISM
Eukaryota; Fungi; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
REFERENCE
1 (bases 1 to 1605)

AUTHORS

Freimoser, F.M., Screen, S., Hu, G. and St. Leger, R.J.
EST analysis of genes expressed by the zygomycete pathogen
Conidiobolus coronatus during optimized secretion of proteins
Unpublished (2002)

JOURNAL

Contact: Freimoser F. M.

Department of Entomology

University of Maryland

4112 Plant Sciences Building, College Park, MD 20742, USA

Tel: 301 405 16 13

Fax: 301 314 92 90

Email: ff34@umail.umd.edu.

Location/Qualifiers

1. .1605

/organism="Conidiobolus coronatus"

/mol_type="mRNA"

/strain="ARSEF 512"

/db_xref="taxon:34488"

/clone_lib="Conidiobolus cornatus ARSEF 512"

/notes="Vector: UniZap; Conidiobolus coronatus was grown in
minimal medium supplemented with Manduca sexta cuticle and
peptone for 18 hours. A cDNA library was constructed in
the unidirectional Lambda vector UniZap."

FEATURES

source

Alignment Scores:

Pred. No.: 1.07e-12 Length: 1605
Score: 235.50 Matches: 88
Percent Similarity: 43.40% Conservative: 37
Best Local Similarity: 30.56% Mismatches: 88
Query Match: 10.42% Indels: 75
DB: 13 Gaps: 14

ORIGIN

US-09-985-689A-5 (1-433) x BQ622771 (1-1605)

Qy 6 GlyIleValIysAlaAspValAlaGlnAsnAsn-----TyrGlyLeu----- 19
Db 718 GGTGTTCTAGGCTATTGCGGTCACAGCAATGCTTCCTGGGCTTTGCTGGTGGT 777
Qy 20 -----TyrGlyGlnGly 23
Db 778 CAACGTGCTAAGCTGGATCTGCTTACTTACTTACCAACCATGCTGCTGCTCAAGT 837
Qy 24 GlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHis 43
Db 938 GTCACCTTTTTCGTTTGTAGTACTGGTGTCAATGTCAGCCACCAATGAC----- 885
Qy 44 GluAlaPheArgGlyLysIleThrAlaLeuTyzAlaLeuGlyArgThrAsnAla--- 62
Db 886 -----TTCGGTGTCTGCGCACCT-----TGGGGTACTAACACTGCTGCTGT 924
Qy 63 -----AsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 925 GGTAGCAACACTGATGCTACGGTCACGGTACTCAGTGTGCTGCTGCTGCTGCTGCT 984
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet----- 97
Db 985 ACCTAT-----GGTGTGTCACAGAGCTAACATTTGTCGCTTAAGGCTTTAGTGAT 1038
Qy 98 AspSerSerGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
Db 1039 GATCGCTCCGGATCATCTCTGGAAATTTCTCCGGTATTCAGTGGTGTGAAGCAC--- 1095
Qy 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyr 137
Db 1096 ---TCTGCTCCAGAAAGTTATCTCTATGATTTAGAGGTGTAAGACGATGCTCTT 1152
Qy 138 ThrAlaAsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPhe 157
Db 1153 -----AACACTGCTGTTTAAACAGCGTGTCAAGAGGAGTGTCACTGTGTC--- 1200
Qy 158 AlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsn 177
Db 1201 GCTGTGTTAAGATACAGAGATGCTGTGGTACTCT---CCCGCTTCTGCTCTCTCC 1257

Qy 178 AlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsn 197
Db 1258 GCCATTACCGTTGGTGGCACT-----GATGTC 1284
Qy 198 ProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysPro 217
Db 1285 AATGATAAAAGGCTTCATTCTCTAACTCGTAGTGT-----GTC 1326
Qy 218 AspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer 237
Db 1327 GATATCTTAGTCTCTGCTGTCACATCTCTCCACC----- 1362
Qy 238 SerPheTyrAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThr 257
Db 1363 -----TGGAAGGATCTAACACTGCCACCAACCATCTCTGTTACCTATGCTTGC 1416
Qy 258 ProIleValAlaGlyAsnValAla 265
Db 1417 CCTCACATTGCTGTTTAGTGTCT 1440

RESULT 2

AK029048

LOCUS

DEFINITION

AK029048

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

AUTHORS	TITLE	COMMENT
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katch, H., Kawaji, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saigo, R., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	Direct Submission Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ , Tel: 81-45-503-9222, Fax: 81-45-503-9216]	cdNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/

Alignment Scores: 3 21e-11 Length: 4198
Pred. No.: 226.00 Matches: 114
Score: 226.00
Percent Similarity: 38.55% Conservatives: 56
Best Local Similarity: 25.85% Mismatches: 167
Query Match: 10.08% Indels: 104
DB: 11 Gaps: 21

US-09-985-689A-5 (1-433) x AK029048 (1-4198)

Qy	8	ValLysAlaAspValAlaGlnAAsnAsnTy	GlyLeuTy	GlyGlnGlyGlnValValAla	27
Db	1005	CTCAGCGCAGATGTCGTGGG	---CAGATGGGATACACAGGTGCTTAATGTCAGATGTC	1061	
Qy	28	ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg	47		
Db	1062	GTITTTTGATACGGGCTC	---AGTGAAGAAGCATCCGCAATTTAAG	1103	
Qy	48	GlyLysIleThrAlaLeuTy	AlaLeuGlyArgThrAsnAsnAlaAsn	---63	
Db	1104	AAT	---GTGAAGGAGAGAACCACTGAGCCATGAGGGACCCCTG	1145	
Qy	64	---AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsn	82		
Db	1146	GATGATGGGTAGGCCATGGCACATTCGTTCGAGGTGTCATTCGCCAGCATGAGGGAGTGC	1205		
Qy	83	LysGlyMetAlaProGlnAlaAsnLeu	---ValPheGlnSerIleMetAspSerSerGly	101	
Db	1206	CARGAGATTGCTCCAGATGAGAGCTGCACATCTTCAGGGTCTTTACCAACAATCAG	---1262		
Qy	102	GlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGly	121		
Db	1263	---GTGCTTACACATCTTGGTTCTTGATGCCTTCAACTATGCC	1304		
Qy	122	AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValAlaTy	ThrAlaAsnSer	141	
Db	1305	ATCCTAAAGAGATGAGGTCTCAACCTTAGCATCGGTGGGCCCACTTCATGGATCAT	1364		
Qy	142	ArgGlnValAspGluTy	---ValArgAsnAsnAspMetThrValLeuPheAlaAla	159	
Db	1365	CCGTTTGTTCACAAAGTGTGGAAATTAACAGCTAACATGTAATATATGGTTCTGCTATT	1424		
Qy	160	GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAlaIle	179		
Db	1425	GGCAATGATGGACCTCTCTATGGCATCTGAAATAACCTTCTGCATCAGATGGATGATT	1484		
Qy	180	ThrValGlyAlaThrGluAsnTyArgProSerPheGlySerLeuAlaAspAsnProAsn	199		
Db	1485	GGAGTGGGTGGC	---ATTGACTTTGAAGAT	1511	
Qy	200	HisIleAlaGlnPheSerSerArgGlyAlaThrArg	---AspGly	213	
Db	1512	AACATCGCTCGCTTTCTTCAGGGGAATGACTACTCTGGGAATACAGAGAGCTATGGT	1571		
Qy	214	ArgIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeu	233		
Db	1572	CGTGTGAAGCCTGACATGTGCACTATGCTGGTCTGGAGTGGGGGT	---1616		
Qy	234	AlaProAspSerSerPheTrpAlaAsnTy	AsnSerLysTy	AlaTy	253
Db	1617	---TCCGGTGTGAAGGGGGCTGCCGTGCATCTCTCAGGGACC	1655		
Qy	254	SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLys	273		
Db	1656	AGTGTGCTTCCCGCAGTGTGCGTGGGGCGCTCACCTTGTAGTAAGCACAGATACAGAAG	1715		
Qy	274	AsnArgGlyIleThrProLysProSerLeuIleIleValAlaLeuIleAlaGlyAlaThr	293		
Db	1716	CGGAGGTGGTGG	---AATCTCGCAGTGTGAAGAACAGCTTTGTAGCGTCAGCCCG	1769	
Qy	294	AspValGlyLeuGlyTyProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLys	313		
Db	1770	AGACTCTCT---GGGGTCAACATGTCGACAGAGGTTCATGGCAATGTGATCTGCTCGCA	1826		

```

QY 314 SerLeuAsnValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThr 333
Db 1827 GCTATCAGATCTC-----AGCAGC 1847
QY 334 TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeu----- 348
Db 1848 TATAAACCCGAGCAAGCCTG---AGTCCTAGCTACATCGACCTGACTGAGTGCCCTAC 1904
QY 349 ValTyrThrAspAlaProGlySerThrThrAlaSerTyr-----ThrLeu 363
Db 1905 ATGTGG-----CCCTACTCTCCAGCCCTATCTACTATGGAGGAATGCCAATC 1955
QY 364 ValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAsp 383
Db 1956 GTTAAT-----GTCAACATCTCATATGGCATGGCGCTCACAGGAAGAAAT 2000
QY 384 PheSerTyrProTyrAspAsnAsnTyr-----AspGlyArgAsnAsnVal 398
Db 2001 GTGGATAAGCCT-----GAGTGGCAGCCCTATTTACCACAGNAATGGAGCAACAT 2051
QY 399 GluAsnValPhe-----IleAsnAlaProGlnSerGlyThrTyrThrIleGlu 414
Db 2052 GAATGGCCCTTCCTACTCTCCTAGTGTGTGGCCCTGTGTGAGTTACCTTGCCATCTCC 2111
QY 415 Val 415
Db 2112 ATT 2114

```

```

RESULT 3
LOCUS BJ393752
DEFINITION BJ393752 Dictyostelium discoideum cDNA library, SF Dictyostelium
discoideum cDNA clone dds32b16 5', mRNA sequence.

```

```

ACCESSION BJ393752
VERSION BJ393752.1 GI:19304838
KEYWORDS EST.
SOURCE Dictyostelium discoideum

```

```

ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

```

```

REFERENCE 1 (bases 1 to 594)
AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

```

```

FEATURES
source
1..594
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds32b16"
/sex="mat A"
/dev_stage="Slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

```

```

ORIGIN
Alignment Scores:
Pred. No.: 2,13e-11 Length: 594
Score: 215.50 Matches: 55
Percent Similarity: 53.18% Conservative: 37
Best Local Similarity: 31.79% Mismatches: 60
Query Match: 9.53% Indels: 21
DB: 12 Gaps: 8

```

```

US-09-985-689A-5 (1-433) x BJ393752 (1-594)

```

```

QY 49 LysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspProAsnGlyHis 68
Db 21 AAAGTTGTAATCTATATTACCATCAACAAAGCGAGATAGTGTAAAGTGGATGGTCAC 80
QY 69 GlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn----- 82
Db 81 GGTACACATATTTGTGGTTCTGCAGAGAGTACTCCAGAGGATTTCTTCAGTATATTTCA 140
QY 83 -----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 141 TCATTTAGTGGTCTTGCACACTGATGCAAGATGTCATTTCTTTGATTGGCAAGTGGTCA 200
QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAla 120
Db 201 TCAAGTTTGACACCT---CCATCGAATTTGAACAATTTATCAACATATATATGAGCA 257
QY 121 GlyAlaArgIleHisThrAsnSerTyrGlyAla-----ProValAsnGlyAla 136
Db 258 GGTCGAAGAGTGCATTGTGATTTCTGGGTCTCTGTATCAGTAGAGGGGTATACAGGTAG 317
QY 137 TyrThrAlaAsnSerArgGlnValAspGluTyrVal---ArgAsnAsnAspMetThrVal 155
Db 318 TATTCATCAGACACTGTTCAATTTGATGATTTCTTTCTCACTCATCCAGATTTCACTT 377
QY 156 LeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSer-----AlaProGly 173
Db 378 CTTAGAGCTGCTGTTAAC-----AACGAGCAATACCTATCACTCACTCAATCC 428
QY 174 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySer 193
Db 429 ACTCAAGAAGATGTTATTACCGTTGGTCTCATCAACAATTCATGAAATTAATAT 482
QY 194 LeuAlaAspAsnProAsnHisIleAlaGlnPheSerSer 206
Db 483 TTAACGTGATGCTCCAAATATATAAATATCAATCATCT 521

```

```

RESULT 4

```

```

LOCUS BJ395336
DEFINITION BJ395336 Dictyostelium discoideum cDNA library, SF Dictyostelium
discoideum cDNA clone dds38b16 5', mRNA sequence.

```

```

ACCESSION BJ395336
VERSION BJ395336.1 GI:19306422
KEYWORDS EST.
SOURCE Dictyostelium discoideum

```

```

ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

```

```

REFERENCE 1 (bases 1 to 640)
AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

```

```

FEATURES
source
1..640
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds38b16"
/sex="mat A"
/dev_stage="Slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

```

```

ORIGIN
Alignment Scores:
Pred. No.: 3,06e-11 Length: 640
Score: 214.50 Matches: 67

```

Percent Similarity:	50.47%	Conservative:	41
Best Local Similarity:	31.31%	Mismatches:	65
Query Match:	9.49%	Indels:	41
DB:	12	Gaps:	12

US-09-985-689A-5 (1-433) x BJ395336 (1-640)

QY	19	LeuTyrGlyGlnGlyGlnValValAlaValAlaLalalePThrGlyLeuAspThrGlyArg-----	37
Db	22	TTAAGAGGTAAAGGTTCAGATATTGGATTAATCTCGTACTGTTAGATGTTAGACCATTTGT	81
QY	38	-----AsnAspSer-----SerMetHisGluAlaPheArgGly	48
Db	82	TTCCTTTTCAGATTCGAAGTATCCATACCATTAAATCAAGTGATGAATAATCATAAAAA	141
QY	49	LysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsnAspPro-----Asn	66
Db	142	GTITGAACCTTATAT-----ACTTACCATGACAATGAAGATTATGTAAT	186
QY	67	GlyHisGlyThrHisValAlaGlySerValLeuGlyAsn-----Ala	80
Db	187	GTTCATGTTACACATGTTTGGCTCTCCACAGAGTACTCCAGAGGATTCCTCATGGGCT	246
QY	81	LeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp	98
Db	247	ATTTCATCATTTTAGTGGTCTTGCAACTCACGCCAAGATTCGATTTTATGATCTT-----	300
QY	99	SerSerGlyGlyLeuGlyGlyLeu---ProSerAsnLeuAsnThrLeuPheSerGlnAla	117
Db	301	TCATCTGGAAGTCTTGAACCAACACCCGGAAGATTACAGTCAAATGTACAAACCATTA	360
QY	118	TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAla-----ProVal	133
Db	361	TATGATGCAGGTGCAAGAGTACATGGTGAATCTTGGGGTCTGTATCTTTCGCAAGTTAT	420
QY	134	AsnGlyAlaTyThrAlaAsnSerArgGlnValAspGluTyVal---ArgAsnAsnAsp	152
Db	421	TATGGFGGTATTCGATGATGCTGGTGGTATTCATGCATCTCTATAGTACCAGAA	480
QY	153	MetThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaPro	172
Db	481	TTCCTCTACTAAGAGCTGCTGGTAAT--AACGAGCTATTGGCATCTTTATTAGCTCAA	537
QY	173	GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyArgProSerPheGly	192
Db	538	GCAACAGCTAAAATGCAATTACAGTTGGTGTGAGCAACAGCTCATGTAAATTTATGTG	597
QY	193	Ser-----LeuAlaAspAsnProasn	199
Db	598	TCAGATGCATTCGATATTATGATTTCTCAGATAATGCTAAT	639

BQ142519 2141 bp mRNA linear EST 24-APR-2000
 Contig6 Metarhizium anisopliae sf. acridum ARSEF 324 Metarhizium anisopliae var. acridum cDNA, mRNA sequence.
 BQ142519
 BQ142519.1 GI:20279578
 EST.
 Metarhizium anisopliae var. acridum
 Metarhizium anisopliae var. acridum
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocerales; Clavicipitaceae; mitosporic Clavicipitaceae; Metarhizium.
 1 (bases 1 to 2141)
 RefSeq(F,M., Screen,S., Bags,S., Hu,G. and St. Leger,R.J.)
 EST analysis of genes expressed by two different insect pathogenic fungi during optimized secretion of proteins
 Unpublished (2002)
 Contact: Freilmoser F. M.
 Department of Entomology
 University of Maryland
 4112 Plant Sciences Building, College Park, MD 20742, USA

RESULT 5	BQ142519/c
LOCUS	BQ142519.1
DEFINITION	Contig6 Metarhizium anisopliae sf. acridum ARSEF 324 Metarhizium anisopliae var. acridum cDNA, mRNA sequence.
ACCESSION	BQ142519
VERSION	BQ142519.1
KEYWORDS	EST.
SOURCE	Metarhizium anisopliae var. acridum
ORGANISM	Metarhizium anisopliae var. acridum
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocerales; Clavicipitaceae; mitosporic Clavicipitaceae; Metarhizium.
AUTHORS	RefSeq(F,M., Screen,S., Bags,S., Hu,G. and St. Leger,R.J.)
TITLE	EST analysis of genes expressed by two different insect pathogenic fungi during optimized secretion of proteins
JOURNAL	Unpublished (2002)
COMMENT	Contact: Freilmoser F. M. Department of Entomology University of Maryland 4112 Plant Sciences Building, College Park, MD 20742, USA

liquid nitrogen prior it's storage at -80 C until RNA extraction. Directional cloning with 5' end of cDNA cloned into EcoRI site of pBluescript and 3' end of cDNA cloned into XhoI site of pBluescript (Stratagene, La Jolla, CA)."

ORIGIN

Alignment Scores:
 Pred. No.: 1-26e-09 Length: 718
 Score: 199.50 Matches: 70
 Percent Similarity: 44.20% Conservative: 29
 Best Local Similarity: 31.2% Mismatches: 84
 Query Match: 8.82% Indels: 41
 DB: 12 Gaps: 9

US-09-985-689A-5 (1-433) x B1750157 (1-718)

```

QY 46 PheArgGlyLeuThrAlaLeuTyrrAlaLeuGlyArgThrAsnAsnAlaAsnAspPro 65
DB 9 TTCGAGGGTCGTGCTCAGGCTGTCTACACTGCTTCAGCGGCCAGACGCT---GACACC 65

QY 66 AsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsnLysGlyMet 85
DB 66 AACGTCACGGAACCTCACGTTCTGGCACTATGTCGGAAGACATAC-----GGTGT 119

QY 86 AlaProGlnAlaAsnLeu-----ValPheGln---SerIleMetAspSerSer 100
DB 120 GCCAAGAGAGCCACCACATCCAGCTGTCAAGGTCTTCCAGGGTAGTTCATCCAGCACCTCC 179

QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
DB 180 ATCATCTCTGCTGGCTTCACTGGGCTGCAACGACATCATCTCCAG----- 227

QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrrThralaAsn 140
DB 228 ---GGCCGAACCAAGACACGTCGTCGAATATGTCCTCGGGCGGTGTTACTGCTCTCC 284

QY 141 SerArgGlnValAspGluTyrrValargAsnAsnAspMetThrValLeuPheAlaGly 160
DB 285 TTCACAAACGCTGTCGAGTCTKCTCCAGCTCCGGTATTATCTCGCATTCGTCGGGT 344

QY 161 AsnGlyGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
DB 345 AACGATGGTGCCACAGCTCCACACTTCT---CCTGCTCTCTCCCGCGCCACWCACT 401

QY 181 ValGlyAlaThrGluAsnTyrrArgProSerPheGlySerLeuAlaAsnProAsnHis 200
DB 402 GTCGGTGCC-----ATTGACAGCAACTGGGCC 428

QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
DB 429 ATTGCTCTGACTCTCACTACGTCACGTCCTCTC-----GATATCTTT 470

QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPheTrp 240
DB 471 GCVCCTKGCACCAAGCTTCTCTCGGC-----TGG 500

QY 241 AlaAsnTyrrAsnSerLysTyrrAlaTyrrMetGlyThrSerMetAlaThrProIleVal 260
DB 501 TACACGAGCAACAGTKCCACCAACACCTCAGCGGACGTCCTGCTACTCTCCACAPT 560

QY 261 AlaGlyAsnVal 264
DB 561 GCGGGACTTGTC 572

```

RESULT 8

CA320325 771 bp mRNA linear EST 09-JUL-2003
 LOCUS UI-M-FWO-cby-d-23-0-UI.r1 NIH_BMAP_FWO Mus musculus cDNA clone
 DEFINITION IMAGE:6816072 5', mRNA sequence.
 ACCESSION CA320325
 VERSION CA320325.1 GI:24538449
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 771)

AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgabs@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.

Location/Qualifiers

FEATURES

source

```

1..771
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_FWO"
/notes="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGGACAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., Program coordinator."

```

ORIGIN

Alignment Scores:

Pred. No.: 3.62e-09 Length: 771
 Score: 195.50 Matches: 78
 Percent Similarity: 40.21% Conservative: 37
 Best Local Similarity: 27.27% Mismatches: 114
 Query Match: 8.65% Indels: 57
 DB: 14 Gaps: 10

US-09-985-689A-5 (1-433) x CA320325 (1-771)

```

QY 8 ValLysAlaAspValAlaGlnAsnAsnTyrrGlyLeuTyrrGlyGlnValValAla 27
DB 40 CTGACGACGATGCTGCTGG---CAGATGGGATACAGAGTCTAATGTGACAGTTGCT 96

QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
DB 97 GTTTTTCATCTGGGCTC-----AGTGAGAAGCATCCGCATTTTAAG 138

QY 48 GlyLysIleThrAlaLeuTyrrAlaLeuGlyArgThrAsnAsnAlaAsn----- 63
DB 139 AAT-----GTGAGGAGAGAACCACTGGACCACTGAGCGGACCGCTG 180

QY 64 ---AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsn 82
DB 181 GATGATGGGTAGGCCCATGGCACATCTGTTGCAGGTGTGATTCACCATGAGGAGTGC 240

```

```

Qy 83 LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSerGly 101
Db 241 CAAGGATTGCTCCAGATGAGAGCTGCACATCTTTCAGGCTCTTTTACCAACATCAG--- 297

Qy 102 GlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaThrPheAlaGly 121
Db 298 -----GTGCTTACACATCTGTTCTGATGATCCCTTCACTATGCC 339

Qy 122 AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSer 141
Db 340 ATCTAAAGAAGATGAGACCTTCTCAACCTTAGCATCGTGCGGCCGACCTTCATGGATCAT 399

Qy 142 ArgGlnValAspGluTyr-----ValArgAsnAspMetThrValLeuPheAlaAla 159
Db 400 CCGTTTGTTCACAGGTGTGGAAATTAAACAGCTAACATGTAATATGTTTCTGCTATT 459

Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 460 GGCAATGATGACCTCTCTATGGCACTCTGAATAACCTGCTCATCAGATGGATGTGATT 519

Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
Db 520 GGAGTGGGTGGC-----ATTGACTTTTGAAGAT 546

Qy 200 HisIleAlaGlnPheSerArgGlyValAlaThrArg-----AspGly 213
Db 547 AACATCGCTCGCTTTCTCCAGGGGAATGACTACCTGGGAATTACCAGGAGCTATGGT 606

Qy 214 ArgIleIysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeu 233
Db 607 CGTGTGAAGCTGACATGTGCACCTATGCTGCTGGAGTGGGGGT----- 651

Qy 234 AlaProAspSerSerPheTyrAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThr 253
Db 652 -----TCCGTTGTGAAGGGGGCTGCCGTGCACCTCTCAGGGACC 690

Qy 254 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLys 273
Db 691 AGTGTGCTTCCCGCAGTGGTGGTGGGGCGCTACCTTGTGTAGTANGCACAGTACAGAAG 750

Qy 274 AsnArgGlyIleThrPro 279
Db 751 CGGAGCTGGTGGATCCT 768

RESULT 9
BJ387574
LOCUS
DEFINITION
Dictyostelium discoideum cDNA library, SF Dictyostelium
ACCESSION
Dictyostelium discoideum
VERSION
Dictyostelium discoideum
KEYWORDS
Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
ORGANISM
1 (bases 1 to 601)
Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
Full length cDNA of Dictyostelium discoideum at the slug stage
Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tschini@genes.nig.ac.jp.
Location/Qualifiers
1. 601
/mol_type="mRNA"
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dd83a18"

FEATURES
source

```

```

/sex="mat A"
/dev_stage="slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN

Alignment Scores:
Pred. No.: 3.78e-09 Length: 601
Score: 193.50 Matches: 55
Percent Similarity: 46.34% Conservative: 21
Best Local Similarity: 33.54% Mismatches: 49
Query Match: 8.56% Indels: 39
DB: 12 Gaps: 6

US-09-985-689A-5 (1-433) x BJ387574 (1-601)

Qy 197 AsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLys 216
Db 51 AATGAGAATAATATTGTTTCATCTCAAAAGCTCCAAACACATGATGGTAGAATGAAA 110

Qy 217 ProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAla----- 234
Db 111 CCGTATTAGTTGGCCCTGGTGAATATATACATCGGCAGATCAATGGTCCCAATACA 170

Qy 235 -----ProAspSerSerPheTyrAlaAsnTyrAsnSerLysTyrAlaTyrMet 250
Db 171 ACAGACCAATGTTGGTGGTCTTTA---CCAATACAAATGCTATTATGGCG---ATA 224

Qy 251 GlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHis 270
Db 225 TCTGGTACATCAATGGCACTCTTCATTCGACGACGACCAACAATCTTAGACAATAT 284

Qy 271 -----PheIleLysAsnArgGlyIleThrProLys 280
Db 285 TTAGTTGATGTTATTATCCAACTGGTTCAATTGTAGAAATCAAAATAAATTACAACCACT 344

Qy 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAla----- 292
Db 345 GGATCATTTATTAAAGCATTAATGATTAATATCTCAGTTATTAAATGGTACATTCAA 404

Qy 293 -----ThrAspValGlyLeuGlyTyrProSerGlyAsp----- 303
Db 405 TTGATTACATCATCAAGTATTACATATCCATCAACCAAGTTTGTGAAATTTTGCAGGT 464

Qy 304 -----GlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 465 GCAAGTTTAGTTCAAGTTGGGTGCTATTAGAAATGAGTAATGTTGTTTGTCTCAT 524

Qy 320 ValAsnGluAla 323
Db 525 AATAATAATAGT 536

RESULT 10
BJ387574
LOCUS
DEFINITION
Dictyostelium discoideum cDNA library, SF Dictyostelium
ACCESSION
Dictyostelium discoideum
VERSION
Dictyostelium discoideum
KEYWORDS
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Toxoplasma.
ORGANISM
1 (bases 1 to 545)
Tang,K., Cole,R., Fogarty,S., Sibley,L.D., Ajioke,J.A., White,M.,
Clifton,S., Pape,D., Martin,J., Wylie,T., Dante,M., Marra,M.,
Hillier,L., Kucaba,T., Theising,B., Bowers,F., Gibbons,M.,
Ritter,E., Bennett,J., Franklin,C., Tsagarisshvili,R., Ronko,I.,
Kennedy,S., Maguire,L., Waterston,R. and Wilson,R.
Toxoplasma EST Project
Unpublished (2001)
Contact: Clifton, S.

```

Toxoplasma EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@wustl.edu
Contact David Sibley (toxos@borcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco
High quality sequence stop: 445.
Location/Qualifiers
1. .545

FEATURES

source

```

/organism="Toxoplasma gondii"
/mol_type="mRNA"
/strain="Veg (Type III)"
/db_xref="taxon:S611"
/clone="TGSTZy80b07.y1"
/dev_stage="Partially sporulated oocysts"
/clone_lib="TgVEG Partially sporulated oocyst cDNA"
/note="Vector: Modified pBluescript (pBSSK+); Site 1: BamHI; Site 2: EcoRI; PolyA mRNA from partially-sporulated oocysts was converted to cDNA using the template-switching PCR method (SMART cDNA, Clontech Inc.,) and sized selected on SizeSep 400 columns (Amersham Pharmacia Biotech Inc.). First strand was reverse transcribed using the CDS III-oligo-dT primer and a 5' template switch primer (Smart IV primer). The product of the first strand synthesis was PCR amplified using the same primer set and the fragments were digested with SfiI. The fragments were size selected, ligated into a modified pBluescript vector containing directional SfiI sites, and electroporated into DH10B or DH12S cells. Vector: SfiI sites were added to the multiple cloning region of pBluescript SK+ between the BamHI/EcoRI sites. The modified polylinker has the following sequence:
5'GATTCGGCATTCAGGCC(G)n- insert--
GGCGCTCGGCCACGACCC3' where n=3-4 G nucleotides. Library
Source: Michael White, Maria E. Jerome, Emily A. Johnson, Jay A. Radke, Montana State University. Clone
Availability: David Sibley, Washington University"

```

ORIGIN

Alignment Scores:
Pred. No.: 7,28e-09 Length: 545
Score: 190.00 Matches: 64
Percent Similarity: 44.81% Conservative: 31
Best Local Similarity: 30.19% Mismatches: 75
Query Match: 8.40% Indels: 42
DB: 13 Gaps: 9

US-09-985-689a-5 (1-433) x BUS75479 (1-545)

```

QY 64 AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGly----- 78
Db 4 GACGACACGGCCACGGCAGCATGTGCGCGC---ATCATCGGGCGCATCAAGGACCAA 60
QY 79 AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 98
Db 61 AACTCCCGGTGAAGAGCGCTGTGTGGAAACACAGCATTCGCCGCGTTAAGTTCATGGGT 120
QY 99 SerSerGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrp 118
Db 121 GCAACCGA-----AACGGTCGACGTTCAGATGCCATCAAGGCTCTCAACTACCGCGTC 174
QY 119 AsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrp 138
Db 175 CAGATCGGAATTCCTCGATGCAACTGCTGGCGCGTCCCACTGGTCGAGCGGTG 234
QY 139 AlaAsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAla 158
Db 235 ATTGCGGCTCTGAACGGCGGAGCGTGGTCAC-----CTTTTCATTCGCGG 285
QY 159 AlaGlyAsnGluCluProAsnSerGlyThrIle-----SerAlaProGlyThrAla 175

```

```

Db 286 GCGGGAACCAAGGAAGAACACCGAGAAATTCGCATTACCTCCCTCGTACGGCTC 345
QY 176 LysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAla 195
Db 346 GCAACCTAGTCAGCGTGGCTGCACA----- 372
QY 196 AspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIle 215
Db 373 AACTCCGAGGACCAACTGCTCTCCCTTCAGAACGGGAGCGGCC----- 417
QY 216 LysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPro 235
Db 418 ACGGTTGACCTTCGGCGCGCGCGGCGTTAAATCCTC-----TCTACCTTTCGCCCA 468
QY 236 AspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMet 255
Db 469 GAC-----CAGTTTCGGGAACCTTTCGGGAACCTCCATG 501
QY 256 AlaThrProIleValAlaGlyAsnValAlaGlnLeu 267
Db 502 GCTACGCGCGTGTGCGGGCGTGGCGGCATCCTC 537

RESULT 11
TA315H10P
LOCUS
DEFINITION
T. brucei sheared genomic DNA clone 315h10, forward sequence,
genomic survey sequence.
ACCESSION
AL490202
VERSION
AL490202.1 GI:11866292
KEYWORDS
GSS.
SOURCE
Trypanosoma brucei
ORGANISM
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE
1 (bases 1 to 574)
Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
McIlvaine,S.E., Rajadream,M.A. and Barrell,B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nhl@sanger.ac.uk
COMMENT
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES
source
Location/Qualifiers
1. .574
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="315h10"

```

Alignment Scores:
Pred. No.: 1.27e-08 Length: 574
Score: 188.00 Matches: 51
Percent Similarity: 47.65% Conservative: 30
Best Local Similarity: 30.00% Mismatches: 59
Query Match: 8.31% Indels: 30
DB: 29 Gaps: 6

Pred. No.: 1.43e-08 Length: 614
Score: 188.00 Matches: 68
Percent Similarity: 45.68% Conservative: 43
Best Local Similarity: 27.98% Mismatches: 84
Query Match: 8.31% Indels: 48
DB: 9 Gaps: 12

US-09-985-689A-5 (1-433) x AJ273402 (1-614)

Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 11 GGTGAGGTTACTTCGCTATATATCATCTGACACTGGTATTGAG-----52
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyzAlaLeuGlyArgThrAsn 60
Db 53 GCCTCCACCCCGAGTTGAGGTCGCGCCACTTTCTTAAGAGCTTCATCAGCGGTCAA 112
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 113 AAC---ACTGATGCCACGCCCATGGGACTCCTGCGTGTGTTACCAT---GGTAGC 163
Qy 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer---99
Db 164 AAGACCTACGCTGTTCGCAAAAGGCTTAAGCTTATGCTGTCAGGTTCTTGACACAG 223
Qy 100 ---SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrp 118
Db 224 GGCAGTGGTCTCTACTCCGCTATCATCATGGTGGCATGACTACGTTCGACAGCTCCAAG 283
Qy 119 AsnAlaGly-----AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGly 135
Db 284 ACCCGCGCTGCCCAACCGGCGCATTTGCTTCATGAGCTGGGA-----GGT 331
Qy 136 AlaTyThrAlaAsnSerArgGlnValAspGluTyzValArgAsnAsnMetThrVal 155
Db 332 GGCCTACTCGGCTCCGTCACCAAGGTCGTCTTGTGCTCAATTCGTGTTCTTCTT 391
Qy 156 LeuPheAlaAlaGlyAsnGlyProAsnSerGlyThrIleSerAlaProGlyThrAla 175
Db 392 GCGTTCGCGCTGGCAGCAGTACCGGATGCCGAGAACACCTCT---CCGCTTCGAG 448
Qy 176 LysAsnAlaIleThrValGlyAlaThrGluAsnTyzArgProSerPheGlySerLeuAla 195
Db 449 CCTTCTGCTGCACCTGTTGTGCTCT-----GCG 478
Qy 196 AspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIle 215
Db 479 GAAATGACAGCCGATCTTCTTCTTCAACTAC-----GGCAGAGTT 520
Qy 216 LysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPro 235
Db 521 ---GTCGATATTTTCCTCTCTGAGCAATGTTCTTCCACC-----559
Qy 236 AspSerSerPheTrpAlaAsnTyzAsnSerLysTyzAlaTyzMetGlyThrSerMet 255
Db 560 -----TGG-----ATTGGTGGCGGCACAAACACCATCTCTGTGTACCTNCA 601
Qy 256 AlaThrPro 258
Db 602 GCTACTCCC 610

US-09-985-689A-5 (1-433) x TA315H10P (1-574)

Qy 18 GlyLeuTyzGlyGlnGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 63 GGTATTGACGGTAGTGGTGAATAATAGGGTAGCGGATACGGGTATCGACTTT-----116
Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyzAlaLeuGly 57
Db 117 AACAGCTGTTCTTCCACGATCCA-----AATCAAGAGGTGCGCTTTACCCGAAGTT 170
Qy 58 ArgThrAsnAsn-----61
Db 171 AACTATACACCGCAAAATCGTGTCTATTGCCCCGTGTGACTTCATCCGGGGGATTAC 230
Qy 62 -----AlaAsnAspProAsnGlyHisGlyThrHisVal-----AlaGlySer 75
Db 231 TTTGCTGGGATGAGGAAATAGTTCATGTCAGCAGCAGTGGCAGGTACCCGACGGGGAGT 290
Qy 76 ValLeu-----GlyAsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 291 GTTATTAGTAAACACCGGTATGCAAGATATATGTTGTCGCCAAGGGGCGAAGATTTC 350
Qy 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThr 112
Db 351 TTCAGGGGTTGGCTGCCATCCAGTCAGAGCTTGTCTCCGCCACGAGTTACTCAA 410
Qy 113 LeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro 132
Db 411 ATTATTGCTCCCGATATGCGCTGGAGCCGCTGTCTTCAAACTCGTGGGTTTGT 470
Qy 133 ValAsnGlyAlaTyThrAlaAsnSerArgGlnValAspGluTyzValArgAsn---Asn 151
Db 471 GCTCCCTCCGAGTATTCTGCTGTGGAAAGGATATGATGAGTTTGGAGTAGTTATGAC 530
Qy 152 AspMetThrValLeuPheAlaAlaGlyAsn 161
Db 531 GATCGCTACTTATCTTCCACTGGCAAC 560

RESULT 12
AJ273402 614 bp mRNA linear EST 29-DEC-1999
DEFINITION AJ273402 Metarhizium anisopliae ARSEF 2575 Metarhizium anisopliae
cDNA clone Ma#948, mRNA sequence.
ACCESSION AJ273402.1 GI:6432774
VERSION EST.
KEYWORDS Metarhizium anisopliae
SOURCE Metarhizium anisopliae
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Clavicipitaceae; Clavicipitaceae; mitosporic
Clavicipitaceae; Metarhizium.
REFERENCE 1 (bases 1 to 614)
AUTHORS Screen,S.E., Mathur,P. and St. Leger,R.J.
TITLE EST analysis of the insect pathogenic fungus Metarhizium anisopliae
JOURNAL Unpublished (1999)
COMMENT Contact: Screen SE
Entomology
University of Maryland
4112 Plant Sciences Building, College Park, MD 20742, USA.
FEATURES
source 1..614
/organism="Metarhizium anisopliae"
/mol_type="mRNA"
/strain="ARSEF 2575"
/db_xref="taxon:5530"
/clone_lib="Metarhizium anisopliae ARSEF 2575"
/note="Vector: UniZap; Metarhizium anisopliae was grown on insect cuticle for 24 hours. A cDNA library was constructed in the unidirectional Lambda vector, UniZap"

ORIGIN
Alignment Scores:
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

```

REFERENCE
AUTHORS
1 (bases 1 to 2121)
Askenazi,M., Driggers,E.M., Holtzman,D.A., Norman,T.C., Iverson,S.,
Zimmer,D.P., Boers,M.E., Blomquist,P.R., Martinez,E.J.,
Monreal,A.W., Feibelman,T.P., Mayorga,M.E., Maxon,M.E., Sykes,K.,
Tobin,J., Cordero,E., Salama,S.R., Trueheart,J., Royer,J.C. and
Madden,K.T.
TITLE
Integrating transcriptional and metabolite profiles to direct the
engineering of lovastatin-producing strains
JOURNAL
Unpublished (2002)
COMMENT
Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
FEATURES
Location/Qualifiers
1..2121
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:3178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone
library"
/note="Vector: pZerOTM-2; Site 1: Sau3A; Site 2: BamHI;
Sau3A genomic fragments ligated into BamHI digested
pZerOTM-2"
ORIGIN
Alignment Scores:
Pred. No.: 3,4e-07 Length: 2121
Score: 183.50 Matches: 79
Percent Similarity: 38.97% Conservative: 34
Best Local Similarity: 27.24% Mismatches: 86
Query Match: 8.12% Indels: 92
DB: 28 Gaps: 12
US-09-985-689A-5 (1-433) x BZ424995 (1-2121)
QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1281 GGGAGGGGACCTACCTTACGTCGTAGACACGGGATCAACGTCGAC-----1234
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsn 60
Db 1233 -----CACGAGAAATTCAGGGCCGTCGAGTCTCGCTACACGCTCCGCGTCCGAC 1180
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 1179 CATGTG---GACGGTGCGCCATGCGACCCAGCTTCTGTACATTCGTGTAAGACA 1123
QY 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle-MetAspSerSe 100
Db 1122 TAC-----GGCGTGGCCAAAGAGGCCACTGCTG-----TCGGTCAAGGTCTTCGTC 1075
QY 100 rGlyGlyLeuGlyGlyLeuProSerAsnLeuAsn-----ThrLeu 113
Db 1074 GGGGAATCGAGCAGCACATCCATCTCTGGACGGCTTCACCTGGCGGCCCATGACATT 1015
QY 113 upheSerGln-----AlaTrpAsnAlaGlyAlaAr 123
Db 1014 GTTTCGAAGAAGCGCACTGGAAAGGCGAGCGATCAACATGAGCTGGGTATGTTATTCGC 955
QY 123 gileHisThrAsnSerTrp-----GlyAlaProValAsnGlyAlaTy 137
Db 954 CATCTATGTGAATGA-TATCATTGGAGTCCACAGCTGACGATGCGCAGCATGCGCGGATA 996
QY 137 rThrAlaAsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPh 157
Db 895 CTCCAAGCCCTCAATGATCGCGTGGAGAGCGCTTTCACAGGAGGATCGTCCCATCGT 836
QY 157 eAlaAlaGlyAsnGluGly-----ProAsnSerGlyThrIleSer-- 170

```

```

Db 835 CGCGCGCGCATGAGATGTGCTGCCACGTTCCACCTGTGAATGGAGTCCGACTCC 776
QY 171 -----AlaProGly-ThrAlaLysAsnAlaI 179
Db 775 GCTGACTGATTCGACAGACCGAGCGCTCGAGCACCGCCGCTTCTGCTCTGATGCC 716
QY 179 leThrValGlyAla-----ThrGluAsnTyArgProSerPheGlySerI 194
Db 715 TTACCGTCTCGGATCAACGTAACACACCGCTGCTATTCTCCAACTACGGGTCCG 656
QY 194 euAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrAgAspGlyA 214
Db 655 TGGTG----- 651
QY 214 rGileLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuA 234
Db 650 -----GATATCTTCGCCCCGGCCAGACATCTCTCTGCC----- 615
QY 234 laProAspSerSerPheTrpAlaAsnTyArgSerLysTyAlaTyMetGlyGlyThrS 254
Db 614 -----TGGATCGGCTCCACACGCGCCACCAACCATCTCGGGCACCT 572
QY 254 erVerAlaThrProIleValAlaGly 262
Db 571 CCATGGCCACCCCCACATGTCGCG 546
RESULT 14
LOCUS
BQ770462 665 bp mRNA linear EST 26-JUL-2002
DEFINITION
UT-M-F10-byV-m-19-0-UI.r1 NIH BMAP_F10 Mus musculus cDNA clone
IMAGE:5702970 5', mRNA sequence.
ACCESSION
BQ770462 GI:21978936
VERSION
BQ770462.1
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 665)
REFERENCE
NTH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rcaphe-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1..665
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5702970"
/tissue_type="whole brain"
/dev_stage="embryo 12.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP_F10"
/note="Organ: Brain; Vector: pYX-Asc; Site 1: Ecor I;
Site 2: Not I; The library was constructed according
to Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated

```

with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CACCACGAC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): "Gene Discovery in the Developing Mouse Nervous System", supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

Alignment Scores:
Pred. No.: 5.93e-08 Length: 665
Score: 182.50 Matches: 65
Percent Similarity: 42.68% Conservative: 37
Best Local Similarity: 27.20% Mismatches: 96
Query Match: 8.07% Indels: 41
DB: 13 Gaps: 7

US-09-985-689A-5 (1-433) x BQ770462 (1-665)

```
QY 63 AsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsn 82
DB 14 GATGATGGGTAGGCCATGCGACATTCGTTGACAGGTGTGATCCACGATGAGGGAGTGC 73
QY 83 LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSerGly 101
DB 74 CAGGATGTCCTCAGATGACAGCTCCATCTTCAGGGCTTTTACCACATCAG--- 130
QY 102 GlyLeuGlyGlyLeuProSerAsnLeuAsnThrPheSerGlnAlaThrPheAsnAlaGly 121
DB 131 -----GTGCTTACATCATCTGTTCTGATGCGCTTCACTATGCG 172
QY 122 AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaThrAlaAsnSer 141
DB 173 ATCTTAAGAGATGAGCTTCTCACTTAGCATCGGTGGCCGCGACTTCATGATCAT 232
QY 142 ArgGlnValAspGluTyr-----ValArgAsnAsnAspMetThrValLeuPheAlaAla 159
DB 233 CCGTTTGTGACAGGTGTGGGATTAACAGCTAACATGTAATGTTGCTGCTATT 292
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
DB 293 GCGAATGATGACCTCTCTATGGCAGCTCTGAATTAACCTCGTATCAGATGGATGATT 352
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
DB 353 CGAGTGGGTGCG-----ATTGACTTGAGAT 379
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArg-----AspGly 213
DB 380 AACATCGCTCGCTTTCTTCCAGGGGAATGACTACCTGGGAATTACCAGGAGGCTATGCT 439
QY 214 ArgIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeu 233
DB 440 CCGTGGAGCCCTGACATTTGCTACCTATGCTGCTGAGTGGCGGGT----- 484
QY 234 AlaProAspSerPheThrAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThr 253
DB 485 -----TCCGGTGTGAAAGGGGGCTCGCTGCTGACCTCTCAGGGACC 523
QY 254 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLys 273
DB 524 AGTGTCGCTTCCAGTGGTGTGCTGGGCGCTACCTCTGTTAGTAAAGCAGATACAGAG 583
QY 274 AsnArgGlyIleThrProLysProSerLeuIleLysAlaAlaLeuIleAlaGlyAla 292
DB 584 CCGGAGCTGGTG-----AATCCTCCAGTGTGAAGCAGCGCTTTGATAGCTGACGCC 634
```

RESULT 15

BZ893395/c
LOCUS
DEFINITION
BZ893395 716 bp DNA linear GSS 30-JUL-2003
HL2.0177 H1 pUC18 Library Halorubrum lacusprofundi genomic 5',
genomic survey sequence.

ACCESSION

BZ893395 GI:33343985

VERSION

GSS.

KEYWORDS

Halorubrum lacusprofundi

SOURCE

Halorubrum lacusprofundi

ORGANISM

Archaea; Euryarchaeota; Halobacteria; Halobacteriales; Halobacteriaceae; Halorubrum.

REFERENCE

1 (bases 1 to 716)

AUTHORS

Goo, Y., Roach, J., Glusman, G., Baliga, N.S., Deutsch, K., Pan, M., Dasarma, S., Ng, W.V. and Hood, L.

TITLE

Low-pass Sequencing for Microbial Comparative Genomics

JOURNAL

Unpublished (2003)

COMMENT

Contact: Goo Y

Institute for Systems Biology

1441 North 34th Street, Seattle, WA 98103, USA

Tel: 206 732 1412

Fax: 206 732 1299

Email: ygoosystemsbiology.org

Seg primer: M13 Forward

Class: shotgun.

FEATURES

Location/Qualifiers

1..716

/organism="Halorubrum lacusprofundi"

/mol_type="genomic DNA"

/strain="ATCC 49239"

/db_xref="taxon:2247"

/clone_lib="H1 pUC18 Library"

/note="Vector: pUC18, site 1: SmaI; A shotgun library was constructed from Halorubrum lacusprofundi genomic DNA using pUC18/SmaI/BAP plasmid"

ORIGIN

Alignment Scores:

Pred. No.: 6.73e-08 Length: 716

Score: 182.50 Matches: 71

Percent Similarity: 45.45% Conservative: 34

Best Local Similarity: 30.74% Mismatches: 66

Query Match: 8.07% Indels: 61

DB: 28 Gaps: 11

US-09-985-689A-5 (1-433) x BZ893395 (1-716)

QY

60 AsnAlaAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu---Gly 78

DB

701 TCCGATGCTCGACGACGAGCGGACCGAATCAGTCGCGGACCGTTCGCGGGCGGA 642

QY

79 AsnAlaLeuAsnLys-----GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 96

DB

641 AACGCGAGTGGGACCGCAATCGGCTCGCACCGAAGCGTCTCCACGGGATCAAGGTG 582

QY

97 MetAsp-SerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerG1 116

DB

581 TTCGACGACGACGGGA-----CCAAACGGACGCTTCGTCGGGTGCTCGCG 537

QY

116 nAlaTrpAsnAlaGlyAlaArgIleHisThrAsnSerTrp-----GlyAlaProVa 133

DB

536 GAATGGACACGCGACGCGAGGATCCGACGCTCGACGCTTCAGATCAGCTCGTCCG 477

QY

133 lAsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGluTyrValArgAsnAspMe 153

DB

476 GACGGGCACTTACCATCTTCTATCATGACCG-GTT---CGGAACACTCGAGTCCGGGAA 421

QY

153 tThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProG1 173

DB

420 GATCGCGCTGCTTTCGCGCGGAAATCATCGGT-----CAAGGAACGTCGAGTCTCCCG 367

QY

173 yThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySe 193

DB

366 GAACGCTACGATCTCGCTCGCGTCGGAGCGGTCAACGACGCCG----- 321

QY

193 rLeuAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThr----- 210

DB

320 -----GGGCTCGCCGCACTTCTCCAGC---GGAGACCGATCAACAC 283

```
QY 211 -----ArgAspGlyArgIleLysProAs 218
Db 282 GTCAGCGCGTGGGAAGTATGCCCGCGGACTGGCCCGACGAGTACGTGGTTCGGGA 223
QY 218 pValThrAlaProGlyThrPheIleLeuSerAlaArg-----SerSerLeuAlaPr 235
Db 222 CGTGTGGCGCGCAGCGGTGAGCGTATCTCGCGGACCGGGCGGTACAAACATCCGAAA 163
QY 235 oAspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMe 255
Db 162 GGAC-----GGCACCTCCAT 148
QY 255 tAlaThrProIleValAlaGlyAsnValAla 265
Db 147 GCGCGCACCGCACGTCAGCGCGTCCGCGCG 117
```

Search completed: April 4, 2004, 11:55:04
Job time : 2283.58 secs

/strain="VEG (Type III)"
/db_xref="taxon:5811"
/clone="IGSRzy98007.y1"
/dev_stage="partially sporulated oocysts"
/clone_lib="TVEG partially sporulated oocyst cDNA"
/notes="vector: Modified pBluescript (pBSK+); Site 1:
BamHI; Site 2: EcoRI; PolyA mRNA from partially-sporulated
oocysts was converted to cDNA using the
template-switching PCR method (SMART cDNA, Clontech Inc.,)
and sized selected on SizeSep 400 columns (Amersham
Pharmacia Biotech Inc.). First strand was reverse
transcribed using the CDS III-oligo-dT primer and a 5'
template switch primer (Smart IV primer). The product of
the first strand synthesis was PCR amplified using the
same primer set and the fragments were digested with SfiI.
The fragments were size selected, ligated into a modified
pBluescript vector containing directional SfiI sites, and
electroporated into DH10B or DH12S cells. Vector: SfiI
sites were added to the multiple cloning region of
pBluescript SK+ between the BamHI/EcoRI sites. The
modified polylinker has the following sequence:
5'GAATTCGGCCATTACGGCC(G)n-- insert--
GGCGCTCGCCGACGATCC3' where n=3-4 G nucleotides. Library
Source: Michael White, Maria E. Jerome, Emily A. Johnson,
Jay A. Radke, Montana State University. Clone
Availability: David Sibley, Washington University"

ORIGIN

Alignment Scores:
Pred. No.: 5,24e-09 Length: 545
Score: 190.00 Matches: 64
Percent Similarity: 44.81% Conservativeness: 31
Best Local Similarity: 30.19% Mismatches: 75
Query Match: 8.41% Indels: 42
DB: 13 Gaps: 9

US-09-985-689A-4 (1-433) x BU575479 (1-545)

Qy 64 AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGly----- 78
Db 4 GACGACCCAGCGCCAGCGCATGTCGCCGCGC---ATCATCGGGGGGATCAAGGACCA 60
Qy 79 AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 98
Db 61 AACTCCCGGGTGMAGGGCTGTGTGGGAACACGAGCATTCGCCGCTTAAGTTTCATGGGT 120
Qy 99 SerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrp 118
Db 121 GCAACACGA-----AACGGGTGACGTCAGATGCCATCAAGGCTCTCACTACGCCGTC 174
Qy 119 AsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrpThr 138
Db 175 CAGATGGGAATTCCTCAGCTGCACTGCTGGGGCGGTCCACCTGGTCCGAGCGCTG 234
Qy 139 AlaAsnSerArgGlnValAspGluTrpValArgAsnAsnAspMetThrValLeuPheAla 158
Db 235 ATTGGCGGTCTGTGAAGCGCGGAGAGCGTCGCTCAC-----CTTTTCATTGGCGCG 285
Qy 159 AlaGlyAsnGluGlyProAsnSerGlyThrIle-----SerAlaProGlyThrAla 175
Db 286 GCGGGAACACAGGAGAAACACCGCAATTCGCGATTACCTCGCTCTCGTACCGGCTC 345
Qy 176 LysAsnAlaIleThrValGlyAlaThrGluAsnTrpArgProSerPheGlySerIleAla 195
Db 346 GCAACAGTAGTCAGCGTGGCTCGACA----- 372
Qy 196 AspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIle 215
Db 373 AACTCGGAGCAACTCGCTCCCTTCACCAACCGGGGAGCGGCC----- 417
Qy 216 LysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPro 235
Db 418 ACGGTTGACCTTGGCGCGCGCGGTAAATCTCTC-----TCTACCTTTCGCGCA 468

Qy 236 AspSerSerPheTrpAlaAsnTrpAsnSerLysTrpAlaTrpMetGlyGlyThrSerMet 255
Db 469 GAC-----CAGTTTCGCAACTTTCGCGAACCCTCCATG 501
Qy 256 AlaThrProLeuValAlaGlyAsnValAlaGlnLeu 267
Db 502 GCTAGCCCGGTGGTGGCGGGCGTCCGCCCATCCTC 537

RESULT 12

AJ273402 AJ273402 614 bp mRNA linear EST 29-DEC-1999
LOCUS AJ273402 Metarhizium anisopliae ARSEF 2575 Metarhizium anisopliae
DEFINITION CDNA clone Ma#948, mRNA sequence.

ACCESSION AJ273402
VERSION AJ273402.1 GI:6432774
KEYWORDS EST.
SOURCE Metarhizium anisopliae
ORGANISM Metarhizium anisopliae

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocryomycetidae; Hypocreales; Clavicipitaceae; Mitosporic
Clavicipitaceae; Metarhizium.

REFERENCE 1 (bases 1 to 614)
AUTHORS Screen,S.B., Mathur,P. and St. Leger,R.J.
TITLE EST analysis of the insect pathogenic fungus Metarhizium anisopliae
JOURNAL Unpublished (1999)
COMMENT Contact: Screen SE

Entomology
University of Maryland

4112 Plant Sciences Building, College Park, MD 20742, USA.

Location/Qualifiers

1. 614

FEATURES

Source

/organism="Metarhizium anisopliae"

/mol_type="mRNA"

/strain="ARSEF 2575"

/db_xref="taxon:5530"

/clone="Ma#948"

/note="Vector: Unizap; Metarhizium anisopliae was grown on

insect cuticle for 24 hours. A cDNA library was

constructed in the unidirectional Lambda vector, UniZap"

ORIGIN

Alignment Scores:
Pred. No.: 8,14e-09 Length: 614
Score: 189.00 Matches: 68
Percent Similarity: 46.09% Conservativeness: 44
Best Local Similarity: 27.98% Mismatches: 83
Query Match: 8.36% Indels: 48
DB: 9 Gaps: 12

US-09-985-689A-4 (1-433) x AJ273402 (1-614)

Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 11 GGTACGGGTACTGCGTATATATCATTCATGACACTGGTATTGAG----- 52
Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyAlaLeuGlyArgThrAsn 60
Db 53 GCCTCCACCCCGAGTTTGAGGGTCCGCCCTTTTCTTAAGAGCTTCATCAGCGGTCAA 112
Qy 61 AsnAlaSerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 113 AAC---ACTGATGGCCAGCCCATGGGACTCACTGCGCTGGTACCATT-----GGTAGC 163
Qy 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer--- 99
Db 164 AAGACCTACCGTGTGCCAAAAGGCTCTATGGTGTCAAGGTCTTTCGACCAACAG 223
Qy 100 ---SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrp 118
Db 224 GGCAGTGGTTTCTACTCCGGGTATCATCAGTGGCATGACTACGTTGCACAGGACTCCAAG 293

Qy	28	ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg	47
Db	97	GTITTTGATACTACGGGCTC-----AGTGAGAAGCATCGCATTTTAAAG	138
Qy	48	GlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn-----AsnAla	62
Db	139	AAT-----GTGAGGAGAGAACCACCTGGACCATGAGCGGACCCTG	180
Qy	63	SerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsn	82
Db	181	GATGATGGGCTAGGCATGGCACATTCGTCAGGTGTGATTCCAGCATGAGGGAGTCC	240
Qy	83	LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSerGly	101
Db	241	CAAGGATTTGGCTCCAGATCGCAGCTGCACATCTCAGGGTCTTTACCAACATCAG--	297
Qy	102	GlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGly	121
Db	298	-----GTGCTCTACACATCTTGGTTTCTGAGTCGCCCTCACTATGCC	339
Qy	122	AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSer	141
Db	340	ATCCTAAGAAGATGGAGGTTCTCAACCTTAGCATCGTGGGCCCGCACTTCATGGATCAT	399
Qy	142	ArgGlnValAspGluTyr-----ValArgAsnAsnAspMetThrValLeuPheAlaAla	159
Db	400	CGGTTGGTGGACAAGGTGGGAATTAACAGCTAACCAATGTATTATGGTTTCTGCTATT	459
Qy	160	GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle	179
Db	460	GGCAATGATGGACCTCTCTATGGCACTCTGGAATACCCCTGCTGATCAGATGGATGTATT	519
Qy	180	ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn	199
Db	520	GGAGTGGGTGGC-----ATTGACTTTGAAGAT	546
Qy	200	HisIleAlaGlnPheSerSerArgGlyValaThrArg-----AspGly	213
Db	547	AACATCGCTCGCTTTCTTCCAGGGAATGACTACCTGGGAATTACCAGAGGCTATGTT	606
Qy	214	ArgIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeu	233
Db	607	CGTGTCAAGCCTGCATGTCACCTATGGTCTCGAGTGGGGGT-----	651
Qy	234	AlaProAspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThr	253
Db	652	-----TCCGGTGTGAAAGGGGGCTGCGGTGCACCTCTCAGGGACC	690
Qy	254	SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLys	273
Db	691	AGTGTGCTTCCCCAGTGGTGGCTGGGGGGCTCACCTTGTGTAGTANGCACAGTACAGAG	750
Qy	274	AsnArgGlyIleThrPro	279
Db	751	CGGGAGCTGGTGAATCCT	768

RESULT 9	BZ424995	2121 bp	DNA	linear	GSS 13-DEC-2002
BZ424995/c	100023066-5726	<i>Aspergillus terreus</i>	random genomic	DNA clone library	
LOCUS	BZ424995	<i>Aspergillus terreus</i>	genomic,	genomic survey sequence.	
DEFINITION	BZ424995	<i>Aspergillus terreus</i>	genomic,	genomic survey sequence.	
ACCESSION	BZ424995.1	GI:26666450			
VERSION	GSS.				
KEYWORDS	<i>Aspergillus terreus</i>				
SOURCE	<i>Aspergillus terreus</i>				
ORGANISM	<i>Aspergillus terreus</i>				
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;				
AUTHORS	Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; <i>Aspergillus</i> .				
	(bases 1 to 2121)				
	Askenzi,M., Driggers,E.M., Holtzman,D.A., Norman,T.C., Iverson,S.,				
	Zimmer,D.P., Boers,M-E., Blomquist,P.R., Martinez,E.J.,				
	Monreal,A.W., Feilbman,T.P., Mayorga,M.E., Maxon,M.E., Sykes,K.,				
	Tobin,J., Cordero,E., Salama,S.R., Trueheart,J., Rover,J.C. and				

Madden, K.T.
integrating transcriptional and metabolite profiles to direct the
engineering of lovastatin-producing strains
JOURNAL
Unpublished (2002)
COMMENT
Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
Location/Qualifiers
1. .2121
/organism="Aspergillus terreus"
/mol type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db xref="taxon:33178"
/lab hosts="Escherichia coli"
/clone lib="Aspergillus terreus random genomic DNA clone
library"
/note="Vector: pZEROTM-2; Site1: Sau3A; Site2: BamHI;
Sau3A genomic fragments ligated into BamHI digested
pZEROTM-2 "

PZL:OJAN 2
 ORIGIN
 Alignment Scores:
 Pred. NO.: 3.36e-08
 Score: 192.00
 Length: 2121
 Matches: 90
 Conservative: 37
 Percent Similarity: 37.80%
 Mismatches: 103
 Best Local Similarity: 26.79%
 Indels: 107
 Query Match: 8.50%
 DB: 28
 Caps: 15
 US-09-985-689A-4 (1-433) X B242499S (1-2121)

```

US-09-985-689A-4 (1-433) x BZ424995 (1-2121)

Qy 21 GlyGlnGlyGlnValValAlaValAlaAAspThrGlyLeuAAspThrGlyArgAsnAspSer 40
Db 1281 GGCAGGGCACCTACGCTTACGTCGTAGACACCGGCATCAACGTGGAC-----1234
Qy 41 SerMethHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1233 -----CACGAGGAATTGAGGCGCGGCGAGGTCCTCGGTAC-----1198
Qy 61 AsnAlaSer-----AspProAsnGlyHisGlyThrHisValAlaGlySer 75
Db 1197 AACGCTCCGGTGGCCAGCATGTGGACGGTGTGGGCATGCCACCCACGTTCTGGGTACA 1138
Qy 76 ValLeuGlyAsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSer 95
Db 1137 ATTGGTGGTAAAGACATAC-----GGCGTGGCCAAAGAGGCCAACCTGCTG-----TCG 1090
Qy 96 Ile-MetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsn-----111
Db 1089 GTCAAGGTCCTCGTCGGGGAAATCGAGGCACACATCCATCATCTCTGGACGGCTTCAACTGG 1030
Qy 112 -----ThrLeuPheSerGln-----AlaTrp 118
Db 1029 GCGGCCAATGACATGTGTTTCCAAGACCGCACTGGAAAGCGAGTCATCAACATGAGCTTG 970
Qy 118 pAsnAlaGlyValAlaArgIleHisThrAsnSerTrp-----GlyAlaLpr 132
Db 969 GGTATGTTATCTGCCATCTATGTGAATGA-TATCATTTGGAGTCACACAGCTGACGCAAG 911
Qy 132 oValAsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGluTyrValArgAsnAsnAs 152
Db 910 CATAGGCGGTGGATPACTCCAAGGCGTTCAATGATGCCGTCGAGAACCGCTTTCACAGAGGG 851
Qy 152 pMetThrValLeuPheAlaAlaGlyAsnGluGly-----ProAsnSer 166
Db 850 AGTCCTGTCCATCGTCGCGCGGCATGAGAAATGCGGTGCCACGTTCCACCCCTGTGAA 791
Qy 166 rGlyThrIleSer-----AlaProGly-T 174
Db 1111

```

```

Db          363  GTGATCCGCAATGTGCCCGCTGC 340
              |||      |||      |||
RESULT 8
CA320325
LOCUS      CA320325
DEFINITION linear EST 09-JUL-2003
IMAGE:6616072 5', mRNA sequence.
UTR-N-FWO-cby-d-23-0-UI.x1 NIH-BMAP_FWO Mus musculus cDNA clone
CA320325 771 bp mRNA

```

23 2b

40 CTGCAGGCAGATGTGCTGTGG---

||||| TTGCT 96

```

DEFINITION      Fg02_10g08_R_Fg02_AAPC_ECORC_Fusarium_graminearum_mycelium
Gibberella zeae cDNA clone Fg02_10g08, mRNA sequence.
ACCESSION      BI750157
VERSION        BI750157.1
KEYWORDS       EST.
SOURCE         Gibberella zeae
ORGANISM       Gibberella zeae
REFERENCE      Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
AUTHORS       Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
              1 (bases 1 to 718)
              Harris, L.J., Glassco, T., Rocheleau, H., Allard, S., Chapados, J.,
              Couroux, P., De Moors, A., Hattori, J.I., Ouellet, F., Robert, L.S.,
              Singh, J.A., Spott, D. and Tinker, N.A.
              Expressed Sequence Tags from Fusarium graminearum mycelium
              Unpublished (2001)
              Contact: Harris, Linda J.
              Eastern Cereal and Oilseed Research Centre
              Agriculture and Agri-Food Canada
              Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
              CANADA
              Tel: (613) 759-1314
              Fax: (613) 759-6566
              Email: harrielj@em.agr.ca.
FEATURES       Location/Qualifiers
               1..718
               /organism="Gibberella zeae"
               /mol_type="mRNA"
               /strain="DAOM 180378"
               /db_xref="taxon:5518"
               /clone="Fg02_10g08"
               /tissue_type="Mycelial tissue"
               /dev_stage="Assexual"
               /lab_hosts="E. coli (Sure cells)"
               /clone_lib="Fg02_AAPC_ECORC_Fusarium_graminearum_mycelium"
               /notes="Vector: Bluescript SK+/XhoI-EcoRI; Site 1: EcoRI;
               Site 2: XhoI; Mycelial tissue was collected from V8 agar
               plates after a growth period of 6-7 days at 25 C with 14
               hrs (E+/UV) day light exposure. Mycelia was ground in
               liquid nitrogen prior it's storage at -80 C until RNA
               extraction. Directional cloning with 5' end of cDNA cloned
               into EcoRI site of pBluescript and 3' end of cDNA cloned
               into XhoI site of pBluescript (Stratagene, La Jolla, CA)."
ORIGIN
Alignment Scores:
Pred. No.:      7,9e-10      Length:      718
Score:          200.00      Matches:      82
Percent Similarity: 41.87%      Conservative: 39
Best Local Similarity: 28.37%      Mismatches: 110
Query Match:     8.85%      Indels:      58
DBs:            12      Gaps:        12

US-09-985-689A-4 (1-433) x BI750157 (1-718)
QY      46 PheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaSerAspPro 65
DB      9 TTCGAGGTCGCTGCTCAGCTGTCTACACTCCCTTCAGCGGCCGCAAGCGT--GACACC 65
QY      66 AsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsnLysGlyMet 85
DB      66 AACGTCACGGAACACGTCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 119
QY      86 AlaProGlnAlaAsnLeu-----ValPheGln---SerIleMetAspSerSer 100
DB      120 GCCAAGAGGCCACCATCCAGCTGCAAGCTGCAAGCTGCTCCAGGTTAGTTCATCCAGCACCTCC 179
QY      101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
DB      180 ATCATCCTCGCTGGCTTCACTGGGTCGCAACGACATCATCTCCAG----- 227
QY      121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrAlaAsn 140
DB      228 ---GGCCGAACCAAGACACTCAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 284

```

```

QY      141 SerArgGlnValAspGlyTrpValArgAsnAspMetThrValLeuPheAlaAlaGly 160
DB      285 TTCAACAACGCTGTCGAGTCTKVCCTCAGCTCCGGTATTATCTCTGCAATTCGTCGGGT 344
QY      161 AsnGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
DB      345 AACGATGGTCCACGCTGCCAACACTTCT--CTGCTCTTCTTCTCCAGCGGCACWCACT 401
QY      181 ValGlyAlaThrGluAsnTyArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
DB      402 GTCGGTGC-----ATTGACAGCAACTGGGCC 428
QY      201 IleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
DB      429 ATTTCCTCGTACTCCAACTACGTCACGCTCTC-----GATATCTTT 470
QY      221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
DB      471 GCYCTKGCACCAAGGTTCTCTCGGC-----TGG 500
QY      241 AlaAsnTyArgSerLysTyAlaTyMetGlyThrSerMetAlaThrProIleVal 260
DB      501 TACACCAACCAACAGTCCACCAACCATCAGCGKACSTCCATGGCTACTTCYCCACATT 560
QY      261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
DB      561 GCCGAGCTGTCCKC-----TACGGTATTCTGTCAAGGGTGTCTCT--GGT 605
QY      281 ProSerLeuLysAlaAlaLeuAlaGlyAlaThrAspValGlyLeuGlyTyPro 300
DB      606 GTCCTTGTGTTCACCACTGGCTGACCTCTACTGCCACCTCC----- 647
QY      301 AsnGlyAspGlnGlyTrpGlyArgValThrLeuAsnLysSerLeuAsnValAlaTyVal 320
DB      648 -----GGCAAGAWCACTGGCACTCCGCACTCCCACTCCCACTGATCGG 689
QY      321 AsnGluAlaThrAlaLeuAlaThrGly 329
DB      690 CACAACGGCAGACGCTTCAGTAAGCGT 716

RESULT 7
BQ142519/c
LOCUS      BQ142519
DEFINITION Contig6 Metarhizium anisopliae sf. acridum ARSEF 324 Metarhizium
            anisopliae var. acridum cDNA, mRNA sequence.
ACCESSION  BQ142519.1
VERSION    BQ142519
KEYWORDS   EST.
SOURCE     Metarhizium anisopliae var. acridum
ORGANISM   Metarhizium anisopliae var. acridum
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic
            Clavicipitaceae; Metarhizium.
REFERENCE  1 (bases 1 to 2141)
AUTHORS    Freimoser, F. M., Screen, S., Baca, S., Hu, G. and St. Leger, R. J.
TITLE      EST analysis of genes expressed by two different insect pathogenic
            fungi during optimized secretion of proteins
JOURNAL    Unpublished (2002)
COMMENT     Contact: Freimoser F. M.
            Department of Entomology
            University of Maryland
            4112 Plant Sciences Building, College Park, MD 20742, USA
            Tel: 301 405 16 13
            Fax: 301 314 92 90
            Email: ff34@umail.umd.edu.
            Location/Qualifiers
            1..2141
            /organism="Metarhizium anisopliae var. acridum"
            /mol_type="mRNA"
            /strain="ARSEF 324"
            /db_xref="taxon:92637"
            /clone_lib="Metarhizium anisopliae sf. acridum ARSEF 324"
FEATURES   source

```



```

QY 314 SerLeuAsnValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThr 333
Db 1827 GCTTATCAGATCCTC-----AGCAGC 1847
QY 334 TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeu----- 348
Db 1848 TATAAACCGCAGGCAAGCCTG---AGTCCTAGCTACATCGACCTGAGTGCCTCCTAC 1904
QY 349 ValTrpThrAspAlaProGlySerThrThrAlaSerTyr-----ThrLeu 363
Db 1905 ATGTGG-----CCCTACTGCTCCAGCCTTACTACTATGAGGAATGCCAACATC 1955
QY 364 ValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAsp 383
Db 1956 GTTAAAT-----GTCACTCTCAATGGCATGGCGCTCACAGGAAGAATT 2000
QY 384 PheSerTyrProTyrAspAsnAsnTrp-----AspGlyArgAsnAsnVal 398
Db 2001 GTGGATTAAGCCT-----GAGTGGCGACCTTATTACCACAGAATGGAGCAACATT 2051
QY 399 GluAsnValPhe-----IleAsnAlaProGlnSerGlyThrTyrIleIleGlu 414
Db 2052 GAAGTGGCCTTCTCCTACTCTCAGTGTGTGGCCCTGTCAGGTTACCTTGCCATCTCG 2111
QY 415 Val 415
Db 2112 ATT 2114

RESULT 3
LOCUS BJ393752
DEFINITION BJ393752 Dictyostelium discoideum cDNA library, SF Dictyostelium
discoideum cDNA clone dds32b16 5', mRNA sequence.
ACCESSION BJ393752
VERSION BJ393752.1 GI:19304838
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 594)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
1..594
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds32b16"
/sex="mat A"
/dev_stage="Slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN
Alignment Scores: 2.63e-11 Length: 594
Pred. No.: 213.00 Matches: 57
Score: 53.85% Conservative: 34
Percent Similarity: 33.73% Mismatches: 56
Best Local Similarity: 9.42% Indels: 22
Query Match: 12 Gaps: 9
DB:

US-09-985-689A-4 (1-433) x BJ393752 (1-594)

```

```

QY 54 TyrAlaLeuGlyArgThrAsnAsnAlaSerAspPro---AsnGlyHisGlyThrHisVal 72
Db 33 TATATTACCACATCAACAAGCGACATAGTATAGTAAAGTGGATGTCACGGTACACATATT 92
QY 73 AlaGlySerValLeuGly-----AsnAlaLeuAsn-----LysGly 84
Db 93 TGTGTTCTCGAGCAGGTACTCCAGAGATTCTTCAGTTATATATTATTCATCATTTAGTGGT 152
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGly 104
Db 153 CTTGCAACTGATGCAAGAGATTGTCATTCTTTGATTGGCAAGTGGTTCATCAAGTTTGACA 212
QY 105 GlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaAspIle 124
Db 213 CCT---CATCGGATTGAAACAATTATATCAACATTATATGACGAGGTGCAAGAGTG 269
QY 125 HisThrAsnSerTrpGlyAla-----ProValAsnGlyAlaTyrThrAlaAsn 140
Db 270 CATTGTGATTCTGGGGTTCTGTATCAGTAGAGGGGTATACAGGTAGTTATTATCATCAGAC 329
QY 141 SerArgGlnValAspGluTyrVal---ArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 330 ACTGCTTCAATTGATGATTCTCTTTCACCTCATCCAGATTTTCATCTTTAGAGCTGCT 389
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSer-----AlaProGlyThrAlaLysAsn 177
Db 390 GGTAAAC-----AACGACGAATACCTATCACTACTCACTCAATCCACTGCAAGAAT 440
QY 178 AlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsn 197
Db 441 GTTATTACCGTTGGTGCTCATCAACAATTCATGAAAATTAT-----TTAACTGATGGT 494
QY 198 ProAsnHisIleAlaGlnPheSerSer 206
Db 495 CCAAATTATATAAATTATCAATCATCT 521

RESULT 4
LOCUS BJ395336
DEFINITION BJ395336 Dictyostelium discoideum cDNA library, SF Dictyostelium
discoideum cDNA clone dds38b16 5', mRNA sequence.
ACCESSION BJ395336.1 GI:19306422
VERSION BJ395336.1
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 640)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
1..640
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds38b16"
/sex="mat A"
/dev_stage="Slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN
Alignment Scores: 3.37e-11 Length: 640
Pred. No.: 212.50 Matches: 66
Score:

```

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, P., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kondo, H., Kouda, M., Koya, S., Kurikawa, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, E., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE
JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9221, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>
Location/Qualifiers

FEATURES

source

```
1. 4198
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:4732484M11"
/db_xref="MGI:2391080"
/db_xref="taxon:10090"
/clone="4732484M11"
/tissue_type="skin"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"
417..3575
/notes="unnamed protein product; membrane-bound
transcription factor protease, site 1 (MG|MGI:1927235,
GB|NM_019709, evidence: BLASTN, 99%, match=3782)
putative"
/codon_start=1
/protein_id="BAC26263.1"
/db_xref="GI:26325018"
/translation="MKLVSTWLVVLVLCGRHLGRLGTRALSKAPCPSCSHLTLLK
VFESTVVEYIVAFNGYFTAKARNSEFISSALKSEVENWRIIPRNPSDDQSFDE
VLIQKEKQAGLITLEDHNPNIKRVTQKRVFSLKFAENPIVPCNETRWQKQOSSR
PLKRALSLGSGFHWATGRSSRLRLRAIPROVATLOADVLWQGYTGANVRVAVFD
TGLSEKHPKVKVKTNTNERTLDDGLGHGTFVAGVTASRECGGPAEALHIFR
VTNNQVSTSWFLDAPNYALLKMDVNLNLSIGDPDMVDFVDKVBELTANNVMS
ATNGDGLTGLNPDQMDVIGVGIDFEDNIARFSSRGMTWELPGYGRKFDIV
TYAGVRSVGSGGCRALSGVSASPVVAGVTLVSTVQKRELNPASVQOALIASA
RLRPNFMEQHGKLDLLRAYQILSSYKPSLSPSYIDLTCEPYMPYCSQIYYG
GMPTIVNVTILANGMGTIRIVDKPWRPYLPQNGDNIIEVAFSVLWPWSQYLAISI
SVTKAASWEGIAQGHMITVASPATELHSGAHTSTVKLPKVKIIPTPRSKRVL
WQVQHNLRYPGVFPEDNLRMKNDPLDWDGHDVHTNFRDMYHLSMSGVEVELGAPF
TCDPATQYGLTLIVDSESEYFPREIAKLBDVDNGLSLVIFSDWNTYVMKRVFDE
NTRQWNPDTGANTPALNELLSVNMWFSDDGLIEGFELVIANHDMYASGCSIAKFFE
DGVVITQTFDQGLEVLKQETAVENVPILGLYIPSEGGGRIVLYGDSNCLDSDHQ
KQCFWLLDALLOQYTSYGVTPSLSHSGNRPPSPGAGLAPPERMEGNHLHRYSKYLEA
HLGDPKRPRLPACPHLSWAKPOLNETAPSNLWIKQLLSIDLKVLVLPNFRSNRPQV
RPLSPGSGAWDIPGGMFGRYNOEVGQTIPVFAFLGMAVLAFFVWQISAKSRPKR
RRPRKRPOLAQAHPARTPSV"
```

CDS

```
polyA_signal
4180..4185
/note="putative"
polyA_site
4198
/notes="putative"
ORIGIN
```

Alignment Scores:

Pred. No.:	5,61e-11	Length:	4198
Score:	224.00	Matches:	113
Percent Similarity:	38.32%	Conservative:	56
Best Local Similarity:	25.62%	Mismatches:	168
Query Match:	9.91%	Indels:	104
DB:	11	Gaps:	21

US-09-985-689A-4 (1-433) x AK029048 (1-4198)

Qy	8	VallyyslaaspvalalaglnaenAsenTyrGlyLeuTyrGlyGlnValVala	27
Dd	1005	CTGCAGCGCATGCTGCTGG---CAGATGGGATACACAGTGTCTATGTGCAGAGTTGCT	1061
Qy	28	ValalaaSpThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg	47
Dd	1062	GTTTTTCATACTGGCTC-----AGTGAGAAGCATCCGCATTTTAAG	1103
Qy	48	GlylysileThrAlaLeuTyrAlaLeuGlyArgThrAsn-----AsnAla	62
Dd	1104	AAT-----GTCAAGGAGAGAACCACTGGACCAATAGACGGACCGCTG	1145
Qy	63	SerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsn	82
Dd	1146	GATCATGGGTAGGCCATGGCACATTGTTCCAGGTGTGATGCCACATCAGGGAGTGC	1205
Qy	83	LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSerGly	101
Dd	1206	CAAGGATTGTCAGATGCAGAGTGCACATCTTCAGGGTCTTTACCAACAATCAG--	1262
Qy	102	GlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTatPAsnAlaGly	121
Dd	1263	-----GTGCTTACATCTGTTTCTGGATGCTTCACTATGCTC	1304
Qy	122	AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSer	141
Dd	1305	ATCCTAAAGAAGATGGACGCTTCTCAACATTAGCATGGTGGCGCCGACCTTCATGGATCAT	1364
Qy	142	ArgGlnValAspGluTyr-----ValArgAsnAsnAspMetThrValLeuPheAlaAla	159
Dd	1365	CCGTTTGTGACAGGTGGGAATTAACAGCTAAACAATATTAATTTGTTCTGCTATT	1424
Qy	160	GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle	179
Dd	1425	GGCAATGATGACCTCTCTATGGCACTCTGAATAACCTGCTGATCAGATGGATGTGATT	1484
Qy	180	ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn	199
Dd	1485	GGAGTGGGTGGC-----ATTGACTTTGAAGAT	1511
Qy	200	HisIleAlaGlnPheSerSerArgGlyValaThrArg-----AspGly	213
Dd	1512	AACATCGCTCGCTTTCTTCCAGGGGAATGACTACCTGGGAATTACCAGCAGGCTATGGT	1571
Qy	214	ArgIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeu	233
Dd	1572	CGTGTGAAGCTGACATTCACCTGCTGGTGGAGTGGGGGT-----	1616
Qy	234	AlaProAspSerSerPheTyrAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThr	253
Dd	1617	-----TCGGGTGTAAGAGGGGGCTGCCGTGCCTCTCAGGAGACC	1655
Qy	254	SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLys	273
Dd	1656	AGTGTGCTTCCCAGTGTGCTGGGGCGCTCACCTTGTAGTAAGACACAGTACAGAAG	1715
Qy	274	AsnArgGlyIleThrProLysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThr	293
Dd	1716	CGGAGCTGGTG-----AATCTCGCAGTGTGAGCAAGCTTTGATAGCTCAGCCGG	1769
Qy	294	AspValGlyLeuGlyTyrProAsnGlyAspGlnGlyTrpGlyArgValThrLeuAsnLys	313
Dd	1770	AGACTTCTCT---GGGGTCAACATGTTCCAGCAAGGTTCATGCAAGTTGGATCTCTCGCA	1826

AUTHORS Freimoser, F.M., Screen, S., Hu, G. and St. Leger, R.J.
TITLE EST analysis of genes expressed by the zygomycete pathogen *Conidiobolus coronatus* during optimized secretion of proteins
JOURNAL Unpublished (2002)
COMMENT Contact: Freimoser F.M.
 Department of Entomology
 University of Maryland
 4112 Plant Sciences Building, College Park, MD 20742, USA
 Tel: 301 405 16 13
 Fax: 301 314 92 90
 Email: ff34@umd.edu.

FEATURES Location/Qualifiers
 source
 1..1605
 /organism="Conidiobolus coronatus"
 /mol_type="mRNA"
 /strain="ARSEF 512"
 /db_xref="taxon:34488"
 /clone_lib="Conidiobolus cornatus ARSEF 512"
 /notes="Vector: UniZap; Conidiobolus coronatus was grown in minimal medium supplemented with Manduca sexta cuticle and peptone for 18 hours. A cDNA library was constructed in the unidirectional Lambda vector UniZap."

ORIGIN
 Alignment Scores:
 Pred. No.: 5,6e-13 Length: 1605
 Score: 236.50 Matches: 88
 Percent Similarity: 43.75% Conservative: 38
 Best Local Similarity: 30.56% Mismatches: 87
 Query Match: 10.46% Indels: 75
 DB: 13 Gaps: 14

US-09-985-689A-4 (1-433) x BQ622771 (1-1605)

Qy 6 GlyIleValIysAlaAspValAlaGlnAsnAsn-----TyrGlyLeu----- 19
 Db 718 GGTGTTCTTAAGCTATGCGTCCAGCAAGCAATGCTTGGGGTCTTGGCTGTGGT 777
 Qy 20 -----TyrGlyGlnGly 23
 Db 778 CAACGTGCTAAGCTCGGATGCTCTTACTTTACCAACCATGCTGATGCTGCAAGT 837
 Qy 24 GlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHis 43
 Db 838 GTCACTGTTTCTGTTTATGATCTGCTGCTCAAGTCAAGTCAAGCAATGAC----- 885
 Qy 44 GluAlaPheArgGlyIysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAlaA--- 62
 Db 886 -----ITCGTGTGCTGCGCACT-----TGGGTACTAACAACACTGCTGCT 924
 Qy 63 -----SerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
 Db 925 GGTAGCAACACTGATGCTCAGGTCAGGCTACTCAGTGTGCTGATATGCTGTGTACC 984
 Qy 80 AlaLeuAsnIysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet----- 97
 Db 985 ACCTAT-----GGTGTGCGCAAGAGCTAACATTTGCGGTAAAGGCTTAGGTGAT 1038
 Qy 98 AspSerSerGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
 Db 1039 GATGCTCCGGATCATCTCTGGAATATTCTCCGGTATTGACTGGGTGTGAAGCAC--- 1095
 Qy 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyr 137
 Db 1096 ---TCTGCTGCCAAGAAAGTTATCTCTAGTTTAGGAGTGTGTAAGACGATGCTTT 1152
 Qy 138 ThrAlaAsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPhe 157
 Db 1153 -----AACACTGCTGTTTAAACAGCTGTCAGCAAGGAGTTGTCATGTTGTC--- 1200
 Qy 158 AlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsn 177
 Db 1201 GCTGCTGTACAGTAACAGAGATGCTTGTGGATCTCT---CCCGCTTCTGCTCTTCC 1257

Qy 178 AlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsn 197
 Db 1258 GCCATTACCGTGTGGTGCCT-----GATGTC 1284
 Qy 198 ProAsnHisIleAlaGlnPheSerSerArgGlyValAlaThrArgAspGlyArgIleLysPro 217
 Db 1285 AATGATAAAAGGCTTCACTTCTTAACCTCGGTAGCTGT-----GTC 1326
 Qy 218 AspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer 237
 Db 1327 GATATCTTAGCTCTGCTGTGTCAACATCTCTCCACC----- 1362
 Qy 238 SerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThr 257
 Db 1363 -----TGGAAGGATCTAACACTGCCACCAACACCATCTCTGTACTCTATGGCTTGC 1416
 Qy 258 ProfileValAlaGlyAsnValAla 285
 Db 1417 CCTCACATTGCTGTTTGTAGTGTCT 1440

RESULT 2
 AK029048
 LOCUS
 DEFINITION Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched library, clone:473484M11 product:membrane-bound transcription factor protease, site 1, full insert sequence.
 ACCESSION AK029048
 VERSION AK029048.1 GI:26325017
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 10349636

2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muranatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159

3
 Shibata, K., Itoh, M., Aizawa, K., Nagacka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanuki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muranatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861

4
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)

5
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 4198)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:40:05 ; Search time 2274.58 Seconds
(without alignments)
5684.703 Million cell updates/sec

Title: US-09-985-689A-4
Perfect score: 2260
Sequence: 1 NDVARGIVKADVAQNNGLY.....EVQAYNVSPGQRFSLAIVH 433

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-DB=/cgn2_1/USPTO.spool/US0985689/runat_31032004_161807_4168/app_query.fasta_1.3498
-QS=EST -OPMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPMT=ptc -NORM=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0985689@cgn_1_1_7084@runat_31032004_161807_4168 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estcm:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	236.5	10.5	1605	13	BQ622771 CC_Contig
2	224	9.9	4198	11	AK029048 Mus muscu
3	213	9.4	594	12	BJ333752 BJ333752
4	212.5	9.4	640	12	BJ335336 BJ335336
5	200.5	8.9	532	29	TA319G10P
6	200	8.8	718	12	BI750157 Fg02_1090
7	200	8.8	2141	13	BQ142519 Contig56 M
8	192.5	8.5	771	14	CA320325 UI-M-FW0
9	192	8.5	2121	28	BZ424995 100023066
10	191.5	8.5	601	12	BJ387574 BJ387574
11	190	8.4	545	13	BU575479 TGEStyB8
12	189	8.4	614	9	AJ273402 AJ273402
13	188	8.3	574	29	TA315H10P
14	185.5	8.2	716	28	BZ833395 HU2_0177
15	181.5	8.0	665	13	BQ770462 UI-M-F10
16	181	8.0	650	9	AJ274038 AJ274038
17	180	8.0	530	29	CNS010FO
18	179	7.9	1029	29	CNS071DW
19	178.5	7.9	675	14	CF727824 UI-M-HB0
20	178	7.9	508	28	AO652212 Sheared D
21	178	7.9	633	12	BJ369190
22	177.5	7.9	576	14	CD255943 ScrP691.
23	177.5	7.9	641	12	BU393925 BU393925
24	177.5	7.9	1002	29	CNS06D6B
25	176.5	7.8	594	14	CF846026 pSB036xA
26	176	7.8	583	9	AJ273745 AJ273745
27	176	7.8	583	9	AJ273947 AJ273947
28	176	7.8	593	9	AJ273918 AJ273918
29	176	7.8	601	9	AJ273050 AJ273050
30	176	7.8	601	9	AJ273921 AJ273921
31	173	7.7	610	9	AJ274218 AJ274218
32	170	7.5	609	9	AJ272712 AJ272712
33	170	7.5	681	14	CB690041 CBST-54-B
34	169	7.5	564	28	AO651427 Sheared D
35	168.5	7.5	615	9	AJ274059 AJ274059
36	168	7.4	794	14	CF737198 UI-M-HD0
37	167.5	7.4	937	12	BG246418 602360428
38	167	7.4	781	13	BU612128 UI-M-EW0
39	166.5	7.4	593	9	AJ273903 AJ273903
40	166.5	7.4	596	9	AJ273379 AJ273379
41	166.5	7.4	619	9	AJ273097 AJ273097
42	166	7.3	887	13	BQ879057 AGENCOURT
43	165	7.3	604	9	AJ273185 AJ273185
44	165	7.3	1572	11	AY107161 Zea mays
45	164.5	7.3	604	9	AJ274126 AJ274126

ALIGNMENTS

RESULT 1
BQ622771
LOCUS CC Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
DEFINITION CC Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
ACCESSION BQ622771
VERSION BQ622771.1 GI:21649940
KEYWORDS EST.
SOURCE Conidiobolus coronatus (Delacroixia coronata)
ORGANISM Conidiobolus coronatus
Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
REFERENCE 1 (bases 1 to 1605)

Job time : 7336.25 secs

```

Db      3007  GCCTAC-----GGCGTCGCCAAGAGGCCAAGATCGTAGGCGTCGGCGTCCGGTGTGTAACAAAC 2954
Qy      100  Ser-----GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
Db      2953  TCGGCGCAGGACCAACCCAGTCGTCGCGCGCATCGACTGGTTCGCCCGGAACGCC 2894
Qy      118  TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro-ValAsnGlyAlaTy 137
Db      2893  GT-CAAGCGCGCGTCGCCAACAT-----GTCCCTCGGCGGCGGCCGGA 2850
Qy      137  rThrAlaAsnSerArgGlnValAspGluTyValArgAsn-----AsnAspMetTh 154
Db      2849  CACGGCC-----CTCGACACGGCGGTACGCAACGCCATGGCTCCGGCGTCAC 2802
Qy      154  rValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyTh 174
Db      2801  CTTTCGCGTGGCGCGGCAACGAGTCGACCAACGCTCCACGAGGTCA---CCCGCAG 2745
Qy      174  rAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyArgProSerPheGlySerI 194
Db      2744  CGTCACGGAGGCCATACGGTCGCGCGGACGACCATCGGACGCCAAGCCGCGC--- 2690
Qy      194  eAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyAr 214
Db      2689  -----TACTCCAACTACGGCTCCGTCCTC----- 2666
Qy      214  gIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAl 234
Db      2665  -----GACCTCTTCGCCCGCGGTTCGTCCTACCTCGGCC----- 2630
Qy      234  aProAspSerSerPheTrpAlaAsnTyAsnSerLysTyAlaTyMetGlyGlyThrSe 254
Db      2629  -----TGGAACTCAAGCGACTCGCGGACCAACACCATCTCCGGTACGTC 2586
Qy      254  rMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAs 274
Db      2585  GATGGCACCACCGCACGTGGCGGCGCGCGCGCTC-----CACCTGCCGCCAA 2535
Qy      274  nArgGlyIleThrProLys-----ProSerLeuIleLysAlaAlaLeuIleAlaGl 291
Db      2534  CCCCTCGGCCACCCCGTCCCAGGTCCGCCACGGCGCTGACGTCCGCCGCCACCACCGCGGT 2475
Qy      291  yAlaThrAspValGlyLeuGlyTyProAsnGlyAspGlnGlyTrpGlyArgValThrLe 311
Db      2474  CGTCACCAACCCCGGCACGGGCTCGCCCAAC----- 2444
Qy      311  uAsnLysSerLeuAsnValAlaTyValAsnGluAlaThrAlaLeuAlaThrGlyGlnLy 331
Db      2443  -----CGGTCCTGTACGTCCGCCGGCGGCACACCCCTCCGGCGCGCG 2397
Qy      331  s-----AlaThrTySerPheGlnAlaGlnAlaGlyLysProLeuLy 345
Db      2396  CTTTCGAGAACACCGGTGACTACACGATCAGCGACAACCTCCACGGTCGAGTCCCGGTGAC 2337
Qy      345  sIleSerLeuValThrTrpThrAspAlaProGlySerThrThrAlaSer-----Ty 361
Db      2336  GGTCTCCGGCGTCTCCGGCAACGCGCCCTCGGCCCTCGCGCTCGAGGTCCACATCGTCCA 2277
Qy      361  rThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyValGl 381
Db      2276  CACGTACATCGCGCACCTCCAGGTCCAGTCTGATCCGCCCGGACGCGCACGGGTACAGCT 2217
Qy      381  yAsnAspPheSerTyProTyAspAsnAsnTrpAsnGlyArgAsnAsnValGluAsnVa 401
Db      2216  CAAG-----TCGTACGGCACCGCGGCGGTTCGGACAACATCAACACCGTACTCGGT 2163
Qy      401  lPheIleAsnAlaProGln-----SerGlyThrTyIleIleGluVal 415
Db      2162  G-----AACGCTCTCTCGGAGGCGGCCAACGCGCACGTGGAAACTGCGGGGTG 2117
```

Search completed: April 4, 2004, 23:08:56

```

Qy 311 uAenlySerLeuAsnValAlaTyrrValAsnGluAlaThrAlaLeuAlaThrGlyGlnly 331
Db 57374 -----CGECTCCTGTACGTGGCGGGGCGGACGACACCTTCCTCGGGCCCGCG 57420
Qy 331 s-----AlaThrTyrrSerPheGlnAlaGlnAlaGlyLyseProLeuLy 345
Db 57421 CTTCCGAGAACACCGGTGACTACAGATCAGCGACCAACTCCACGGTCCGTCGGTGAC 57480
Qy 345 sIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----Ty 361
Db 57481 GGTCTCCGGCGTCTCCGGCAACGCGCCCTCCGCGCTCGCGGTAGAGGTCCACATCGTCCA 57540
Qy 361 rThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyrrValG1 381
Db 57541 CAGTACATCGCGACCTCCAGTCCAGTCCAGTGCATCGCCCCCGACGCGGTACACGT 57600
Qy 381 yAsnAspPheSerTyrrProTyrrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnVa 401
Db 57601 CAAAG-----TCGTACGGCACCGGGCGAGTTCGGACAACTCAACACACCGTACTCGGT 57654
Qy 401 lPheIleAsnAlaProGln-----SerGlyThrTyrrIleIleGluVal 415
Db 57655 G-----AACGCTCTCCGAGGGCGGCCAACGCGACGTGGAACTCGCGGTG 57700

RESULT 15
US-10-084-846A-2/c
; Sequence 2, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; PRIORITY FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 2
; LENGTH: 59816
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-2

Alignment Scores:
Pred. No.: 1,44e-20 Length: 59816
Score: 296.00 Matches: 126
Percent Similarity: 42.11% Conservative: 50
Best Local Similarity: 30.14% Mismatches: 153
Query Match: 13.19% Indels: 90
Ds: 15 Gaps: 19

US-09-985-689A-4 (1-433) x US-10-084-846A-2 (1-59816)
Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 3169 GGGCAGGGATGACGGGTACGTTCATCGACACGGGGTCTC-----CGC 3128
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrr---AlaLeuGlyArgThr 59
Db 3127 ATCACCACACGACATTCGGCGCGCGGGCGCTCTACGGGTACGACGCCATCGACACGAC 3068
Qy 60 AsnAsnAlaSerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 3067 AACACGCCACAGACGGCCACGGCCACGGCCACGACGTGGCCGGCAGCGTCCGCGCGCAAC 3008
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99

```


QY 257 rProLysValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIle 277
Db 4131825 CCGCATGTCGGGGCGCCCGCGATCTCAAGCAGCAGCAC----- 4131783
QY 277 eThrProLysProSerLeu-----IleLysAlaAlaLeuIleAlaGlyAlaThrAspVa 295
Db 4131782 ----CCCAACTGGTCGTCGCCGAGCTCAAGGCGCGTGACCGGTCGCCGAAG---- 4131732
QY 295 lGlyLeuGlyTyProAsnGlyAspGlnGlyTyPrgIleValThrLeuAsnLysSerLe 315
Db 4131731 -GGCGGCAAGTACACGCCGCTTCCAGCAGGCTCGGGCGTATCGCGTCGACAGCGCAT 4131673
QY 315 uAsnValAlaTyxValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyxSe 335
Db 4131672 CAAGCATCGGTGATGCCAACCGAACTCGGTGAGCTTCGGCAT-CCAGCATGGCCGC 4131614
QY 335 rPheGlnAlaGlnAlaGlyLysPro 343
Db 4131613 ACACCGACGACAGCGCGTCACCCA 4131589

RESULT 11

US-10-090-624-5
; Sequence 5, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyoza
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 4765
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-5

Alignment Scores:

Pred. No.: 7,71e-24 Length: 4765
Score: 313.00 Matches: 145
Percent Similarity: 35.35% Conservative: 42
Best Local Similarity: 27.41% Mismatches: 180
Query Match: 13.85% Indels: 162
DB: 13 Gaps: 20

US-09-985-689A-4 (1-433) x US-10-090-624-5 (1-4765)

QY 21 GlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1024 GGCATGTTATGATCATGTGATATGTGTGATACGACCTTACACGCTTCCCGGAC--- 1080
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyxAlaLeu----- 56
Db 1081 -----GAAGTTCACCTTGGCCAGTACACAGCTTACTTATGATGTTGCTTTTATG 1131
QY 57 -----GlyArgThrAspAsnAla-----SerAspProAsn----- 66
Db 1132 TACTACTACGCTCTCTCACTACGCTTCCAGAAATAGATCTCAACGAGGAATATGCA 1191
QY 67 -----GlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 1192 GTATTGGGTGGATGGTTCACGGTTCACGAACTCACGTAGCTGGAACCTGTTGCTGGTTAC 1251

QY 80 AlaLeuAsn----- 82
Db 1252 GACAGCAACAATGATGCTTGGGATTTGGCTCAGTATGACTCTGTGTGAATGGGAAGTGTTC 1311
QY 83 -----LysGlyMetAla 86
Db 1312 TCAAGACTCTATGTTGGGATTATACGAAGCTTACCACAGACACCGTGCAGGGTGTGTGCT 1371
QY 87 ProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeu 106
Db 1372 CCAGTGCCCAATAATAGGCAATTAAGAGTCTTCTAGGAGTGATGGA-----CGGGGTAGC 1425
QY 107 ProSerAsnLeuAsnThrLeuPheSerGlnAlaAlaThrAsnAlaGlyAlaArgIleHisThr 126
Db 1426 ATGTGGGATATTATAGAGGTATCATACACGCAACCCATGCTGTGCAGACGTTATAAGC 1485
QY 127 AsnSerTrpGly-----AlaProValAsnGlyAlaTyxThrAlaAsnSerArgGlnVal 144
Db 1486 ATGAGTCTCGGTGGAAATGCTCCATCTAGATGGTACTGATCCAGAAAGCGTTGCTGTG 1545
QY 145 AspGluTyxValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGlyPro 164
Db 1546 GATGAGCTTACGAAAGTACGGTGTTCGTATTCGTATAGCTGCAGGAATGAAGTCTCT 1605
QY 165 AsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr 184
Db 1606 GGCATTAAACATCGTTGGAAGTCTCGTGTTCGAAACAAAGGCAATAACTGTTGGAGCTGCT 1665
QY 185 Glu----- 185
Db 1666 CGAGTGGCCATTACGTTGGAGTATTATGTTTCCAGCAGCTTGGATATCTCTGATTACTAT 1725
QY 186 -----AsnTyxArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGln 203
Db 1726 GGAATTCTATTACTTCCCGCTACACAAACGTT-----AGAATAGCATTC 1770
QY 204 PheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaProGly 223
Db 1771 TTCTCAGCAGAGCCGAGAAATAGATGGTGAATATAAACCCCAATGATGCTGCTCCAGGT 1830
QY 224 ThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsnTyx 243
Db 1831 TACGGAAATTACTCATCCCTCGCATGGATTGGCGAGCTGACTTC----- 1878
QY 244 AsnSerLysTyxAlaTyxMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsn 263
Db 1879 -----ATGCTGGAACTTCGATGGCTACTCCACATGTCCAGCGGTGTC 1920
QY 264 ValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSerLeu 283
Db 1921 GTTGCACTCTCATAAAGCGGG---GCAAAGGCCGAGGGAATATACTACAATCCAGATATA 1977
QY 284 IleLysAlaAlaLeuIleAlaGlyAlaThr-----AspValGlyLeu 297
Db 1978 ATTAAGAAGGTCTTTGAGAGCGGTGCAACCTGGGTTGAGGAGATCCATATATCTGGGCGAG 2037
QY 298 GlyTyxProAsnGlyAspGlnGlyTyxArgValThrLeuAsnLysSerLeuAsnVal 317
Db 2038 AAGTACACTGACCTTGACCAAGGTCATGCTCTGTTAACGTTACCAAGCTCTGGGAATC 2097
QY 318 AlaTyxValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyxSerPheGln 337
Db 2098 -----CTTAAG 2103
QY 338 AlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThr 357
Db 2104 GCTATAACGGCACCACCTCTCCCAATTGTTGATCATTGGGAGAGCAAGTCTCTACAGCGAC 2163
QY 358 ThrAlaSerTyxThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGly--- 376
Db 2164 TTTGGCGGAGTACTTGGGTGTG---GACGTTATAGAGGTCTCTACGCAAGGAAGTCTCTATA 2220

```
QY 104 -GlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaAr 123
Db 1027 GAGTGGCGGCGCCGAGCAGCGCGGCGGCGTCTTCAACCTGAGCTGGCGGCGGCGACACC 1086
QY 123 gileHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrpThrAlaAsnSerArgG1 143
Db 1087 CCGGAGATCGACCCGCTGGAA-GCGAGGTCAACAGCTC----- 1125
QY 143 nValAspGluTyValArgAsnAsnAspMetThrValLeuPhe-----AlaAlaGlyAs 161
Db 1126 -----TCCGAGGAGAGGCGATCTCTTCCGATCGCGCGCGCAA 1166
QY 161 nGluGlyPro--AsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleTh 180
Db 1167 CGAAGCGGAGTTCGGGAGCAGACCATCGCTCCCGGCGCAGCGCGCGCGCTCAC 1226
QY 180 rValGlyAlaThrGluAsnTyArgProSerPheGlySerIleAlaAspAsnProAsnHi 200
Db 1227 CGTCCGCGCC----- 1253
QY 200 sIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValTh 220
Db 1254 GCTGGCGTCTCTTCCAGCGCGCGCGCGCTTCAAGTCAAGCGCGCTCAC 1313
QY 220 rAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTr 240
Db 1314 CGACCCGCGCTCGACATCACCGCGCC-----GCGGCGCGCGCGCTCATCGA 1364
QY 240 pAlaAsnTyAsnSerLys-----TyAlaTyMetGlyThrSerMetAlaTh 257
Db 1365 CCAGGAGGTTCGGCCAGAGCGGCGGCTACCTCACCTCCGCTTACGTCGATCGCGAC 1424
QY 257 rProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyI1 277
Db 1425 CCGCATGTCGCGGCGCGCGCGCTCTCAAGCAGCAGC----- 1467
QY 277 eThrProLysProSerLeu-----IleLysAlaAlaLeuIleAlaGlyAlaThrAspVa 295
Db 1468 -----CCCAACTGCTCTTCCCGAGCTCAAGCGCGGCTGACCGGCTCCGCGAAG- 1518
QY 295 lGlyLeuGlyTyProAsnGlyAspGlnGlyTrpGlyArgValThrLeuAsnLysSerLe 315
Db 1519 -GGCGGAGATGACCGCGTTCAGCAGCGGCTCGGCGCGCTATCGCGCTGCAAGCGCAT 1577
QY 315 uAsnValAlaTyValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTy-Se 335
Db 1578 CAAGCAGTCGTCATCGCCAAACCCGAACCTCGGTGAGCTTCGGCAT-CCAGCAGTGGCGC 1636
QY 335 rPheGlnAlaGlnAlaGlyLysPro 343
Db 1637 ACACGACGACAGCGCGTCAOCCA 1661

RESULT 10
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication NO: US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HAITORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCES: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
```

```
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Alignment Scores:
Pred. No.: 2,21e-20 Length: 9025608
Score: 325.00 Matches: 113
Percent Similarity: 45.56% Conservative: 46
Best Local Similarity: 32.38% Mismatches: 125
Query Match: 14.38% Indels: 67
DB: 12 Gaps: 12
US-09-985-689A-4 (1-433) x US-10-156-761-1 (1-9025608)
QY 18 GlyLeuTyGlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 4132505 GGGTACGACGCGCAAGGCGCTCAAGATCGCGTCTTGACACCGGTGTGCAC----- 4132455
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGly 57
Db 4132454 -----GCGACCCACCGGACCTCAAGACAGGTGGCGGAGTCCAGAACTTCTCC 4132404
QY 58 ArgThrAsnAsnAlaSerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 4132403 GCCCGCGCGGACGCGCGGCGGCTTCCGTCAGGCGCAGCAGCGTCCATCGCGCG 4132344
QY 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 4132343 GGCACCGCGCGCAAGTCCACGCGCAGTACAGGTGTCCGCGCGCGGAGATCTCTC 4132284
QY 93 PheGlnSerIleXetAspSerSerGly-----GlyLeu 103
Db 4132283 AACGCGAAGTCTCTCGACGACACCGGCTCCGCGGAGCTCCGCGCATCTTGGCGGCG 4132224
QY 104 -GlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaAr 123
Db 4132223 GAGTGGCGGCGGAGCAGGCGCGCGCTCGTCACTGAGCCTGGCGGCGGCGACACC 4132164
QY 123 gIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrAlaAsnSerArgG1 143
Db 4132163 CCGGAGATCGACCCGCTGGAA-GCGGAGGTCAACAAGCTC----- 4132125
QY 143 nValAspGluTyValArgAsnAsnAspMetThrValLeuPhe-----AlaAlaGlyAs 161
Db 4132124 -----TCCGAGGAGAGGCGCATCTCTTCCGATCGCGCGCGCAA 4132084
QY 161 nGluGlyPro--AsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleTh 180
Db 4132083 CGAAGCGGAGTTCGGCGAGCAGACCATCGCTCCCGGCGGCGCGCGCGCGCTCAC 4132024
QY 180 rValGlyAlaThrGluAsnTyArgProSerPheGlySerIleAlaAspAsnProAsnHi 200
Db 4132023 CGTCCGCGGCC-----GTGAACGACGACGCAA 4131997
QY 200 sIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValTh 220
Db 4131996 GCTGGCGTCTCTTCCAGCGCGCGCGCGCTTGAGCGGCGGCGCATCAAGCGCGCGTCA 4131937
QY 220 rAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTr 240
Db 4131936 CGCACCGCGGCTGGACATCACCGCGCC-----GCGGCGCGGCGGCGGCTCATCGA 4131886
QY 240 pAlaAsnTyAsnSerLys-----TyAlaTyMetGlyGlyThrSerMetAlaTh 257
Db 4131885 CCAGGAGGTTCGGCCAGAGCGCGGACCGGCTACCTCACCATCTTCCGTACGTGCGGAC 4131826
```


Percent Similarity: 43.33% Conservative: 56
Best Local Similarity: 28.97% Mismatches: 153
Query Match: 14.51% Indels: 68
DB: 14 Gaps: 14

US-09-985-689A-4 (1-433) x US-10-156-761-1 (1-9025608)

QY 8 ValIyAlaAspValAlaGlnAsnAsn-----Tyr 17
Db 6919813 GTCAGGCGGACATGCGCGGAGAGCAAGCGGAGATCGGTACGCGGCGCGGTGGAGCGC 6918872

QY 18 GlyLeuTy-GlyGlnGlnValValAlaAlaAspThrGlyLeuAspThrGlyArg 37
Db 6918873 GGGCTCAGCGGCGACGGCGTACCGTGGCGGTCTCGACACCGGCGTGCACACC----- 6918926

QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysLeuThrAlaLeuTyValAlaLeuGly 57
Db 6918927 -----ACTCAGCCCGACCTCCGCGCGCGGTCTCCGAGCAAGAGCTTCATC 6918974

QY 58 ArgThrAsnAsnAlaSerProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 6918975 GACGGGAGGAGTCCGCGGACCGCAACGCGGACCGACCGTCCACCTCGACCGTCCGCG 6919034

QY 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 6919035 GGCAGCGCGCGCGCTCCGCGGCGAGCGGCGCGGTCTCGCGCGGTCCAGCTCGCC 6919094

QY 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThr 112
Db 6919095 GTCGGCAAGTGTACGAGCAGCAGCGCG-----GGAAGCGAGTCCAGATC 6919142

QY 113 LeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHisThrAsnSerTrp 129
Db 6919143 ATCGCGGCGATGAATGGCGCGGCGGAGCGTGGTCCGAGGATCTCTCGATGAGCCTC 6919202

QY 130 GlyAla---ProValAsnGlyAlaTyThrAlaAsnSerArgGlnValAspGluTyVal 148
Db 6919203 GGATCGAGCGGCGGCGGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919262

QY 149 ArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThr 168
Db 6919263 GAGGAGACCGCGCGCTCTTCGCGTGGCGCGGCGGAGAACACCGGTCGCCCTCC---TCG 6919319

QY 169 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyArg 188
Db 6919320 ATCGGTCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919364

QY 189 ProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGly 208
Db 6919365 -----GTGACTCATCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919406

QY 209 AlaThrArg---AspGlyArgIleLysProAspValThrAlaProGlyThrPheIleLeu 227
Db 6919407 CCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919466

QY 228 SerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnTyAsnSerLysTy 247
Db 6919467 GCGCGCGGCGTCCAGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919505

QY 248 AlaTyMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 267
Db 6919506 ACCTCCATGACGCTAGTGTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919565

QY 268 ArgGluHisPheIleLysAsnArgGlyLeuThrProLysProSerLeuIleLysAlaAla 287
Db 6919566 CCGGAGCAGCACCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919625

QY 288 LeuIleAlaGlyAlaThrAspValGlyLeuGlyTyThrProAsnGlyAspGlnGlyTrpGly 307
Db 6919626 GAGCACTCGAGCGCTCCGATATCATGTTGGGGCG-----GGT 6919664

QY 308 ArgValThrLeuAsnLysSerLeuAsnValAlaTyThrValAsnGluAlaThrAlaLeuAla 327
Db 6919664 -----GAGTTG 6919811

Db 6919665 CCGGTCACTGTGCGGAGCGCGGTGCGCGC-----CGCGTCAACCGGAGCGGCG 6919712

QY 328 ThrGlyGlnLysAlaThrTySerPheGlnAlaGlnAlaGlyLysProLeuLysIleSer 347
Db 6919713 AGCGCGACCTCGCTTCACCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919772

QY 348 LeuValTrpThrAspAlaProGlySerThrThrAlaSerTyThrLeuValAsnAspLeu 367
Db 6919773 GTCACCTACTCCAACTCTCCGACGAGCGGTC-----GAGTTG 6919811

QY 368 AspLeuValIleThr---AlaProAsnGly 376
Db 6919812 AGCTCGCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919841

RESULT 9
US-10-156-761-3306
; Sequence 3306, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3306
; LENGTH: 3417
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3417)
US-10-156-761-3306

Alignment Scores:
Pred. No.: 2,21e-25 Length: 3417
Score: 325.00 Matches: 113
Percent Similarity: 45.56% Conservative: 46
Best Local Similarity: 32.38% Mismatches: 125
Query Match: 14.38% Indels: 67
DB: 12 Gaps: 12

US-09-985-689A-4 (1-433) x US-10-156-761-3306 (1-3417)

QY 18 GlyLeuTyGlyGlnGlnValValAlaAlaAspThrGlyLeuAspThrGlyArg 37
Db 745 GGGTACGAGCGGAGGCGGTCAAGATCGCGTCTGGACACCGGTGTGAC----- 795

QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGly 57
Db 796 -----GGACCCACCGGACCTCAAGGACCGAGTGGCGGAGTCCCAAGAACTCTCTCC 846

QY 58 ArgThrAsnAsnAlaSerAspAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 847 GCGCGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 906

QY 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 907 GGCACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 966

QY 93 PheGlnSerIleMetAspSerSerGly-----GlyLeu 103
Db 967 AACGCAAGGTCTCTCGAGCACACCGGCTCCGCGGAGCAGCTCCGCGGATCTCTGCGCGGCAIG 1026

Db 1174 GGAATTGCGGCA-----CTTATCTTCAGCAAGCCCGCGCAGAGAACCT 1218
 Qy 282 SerLeuIleValAlaLeuIleAlaGlyAlaThrAspVal 295
 Db 1219 GATGAAGTCAACACAGCTGCTTAATGGACGGT---ACCGATTATGGAAAGATCCGATCCA 1275

Qy 296 -----GlyLeuGlyTyrProAsnGlyAspGln 304
 Db 1276 AATGTTACGGTGCAGGGTATACATCAGCCAGAACAA 1311

RESULT 6

US-10-314-657-1
 ; Sequence 1, Application US/10314657
 ; Publication No. US20030175886A1
 ; GENERAL INFORMATION:
 ; APPLICANT: SHEN, Ben
 ; APPLICANT: CHENG, Yi-Qiang
 ; APPLICANT: TANG, Gong-Li
 ; TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
 ; FILE OF INVENTION: Syntheses and Methods of Use
 ; FILE REFERENCE: 054030-0021
 ; CURRENT APPLICATION NUMBER: US/10/314,657
 ; CURRENT FILING DATE: 2002-12-09
 ; PRIOR APPLICATION NUMBER: PCT/US02/08937
 ; PRIOR FILING DATE: 2002-03-22
 ; PRIOR APPLICATION NUMBER: US 60/278,935
 ; PRIOR FILING DATE: 2001-03-26
 ; NUMBER OF SEQ ID NOS: 214
 ; SOFTWARE: Patent in version 3.2
 ; SEQ ID NO 1
 ; LENGTH: 135638
 ; TYPE: DNA
 ; ORGANISM: Streptomyces atroolivaceus
 US-10-314-657-1

Alignment Scores:

Pred. No.: 4,22e-24 Length: 135638
 Score: 334.50 Matches: 137
 Percent Similarity: 42.17% Conservative: 57
 Best Local Similarity: 29.78% Mismatches: 155
 Query Match: 14.80% Indels: 111
 DB: 14 Gaps: 23

US-09-985-689A-4 (1-433) x US-10-314-657-1 (1-135638)

Qy 6 GlyIleValLysAlaAspValAlaGlnAsnAsn----- 16
 Db 8067 GGCAGGGTGAAGCGATCTGCGCAGCTCCACCCCGCAGATCGCGCGCAGAGGTATGG 8126
 Qy 17 ---TyrGlyLeuTyrGlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThr 35
 Db 8127 GCGGAGGGCGACACCGGCGCAGGAGGTGAAGTCTCGATGCTCGACAGCGCGGACACC 8186
 Qy 36 GlyArgAsnAspSerMetHisGluAlaPheArgGlyLysIle-----ThrAlaLeu 53
 Db 8187 -----GACACCCCGGACCTGTCTGGCGCAGGTGTCTGCACAGCGCCAGC 8228
 Qy 54 TyrAlaLeuGlyArgThrAsnAsnAlaSerAspProAsnGlyHisGlyThrHisValAla 73
 Db 8229 TTGCTCCCGCGCGAG---GAGCAGATCCCGAGCTACACGCCACCGCAGCGACGTCGCC 8285
 Qy 74 GlySerValLeuGly-----AsnAlaLeuAsnLysGlyMetAlaProGln 88
 Db 8286 TCGACCATCTCGGCACCGGCGAGCGCCCTCCAGCGCAAGAGCGGGGTGCGCCCTCCGCG 8345
 Qy 89 AlaAsnLeuValPheGlnSerIleMetAspSerGlyGlyLeuGlyGlyLeuProSer 108
 Db 8346 GCCCGGCTGTCGTCGGCAAGGTGTCAACTCCGAGGGC-----AGCGGCGAGGAATCG 8399
 Qy 109 AsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHis 125
 Db 8400 TGGATC-----ATCGCGGGATGGAGTGGGCGCGCCCGCAGCAGAAAGGCCAGGATCATC 8453

RESULT 7

US-10-156-761-5701
 ; Sequence 5701, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN

Qy 126 ThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn-----SerArg 142
 Db 8454 AGCATGAGCCCTGGGC-----GGCGCGGTGACAAAGACGACCCCGATCAGCGAC 8501
 Qy 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
 Db 8502 GCCGTGCGAGCAACTCAGCCACGACGCGCGCGTGTTCGTGATCGCGCGGCAACGCGC 8561
 Qy 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 Db 8562 GSCCGGCACTCC-----ATCAGCAGCCCGGTGCGGCAGACTCCGCGCTACCGTCGGC 8615
 Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
 Db 8616 GCC-----GTGACTCCACCGACCGCTCGCC 8642
 Qy 203 GlnPheSerSerArgGly---AlaThrArgAspGlyArgIleLysProAspValThrAla 221
 Db 8643 GACTTCTCCAGCAGGCGCGCGTGCAGCGGCGCGGTGAAGCGGAGATCACCAGCG 8702
 Qy 222 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAla 241
 Db 8703 CCGCGCGTGCATGCTCGCGCGCGTTCGCACTACAAGCGCGGTCTCGGCTAC----- 8756
 Qy 242 AsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAla 261
 Db 8757 -----TACACCAAGTACGCGCAGCGCTGATGGCGACCGCCACGCTCGCC 8801
 Qy 262 GlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysPro 281
 Db 8802 GGTGTGCGCGCGCTCTCTCGCGCGCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCG 8861
 Qy 282 SerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrProAsn 301
 Db 8862 GCACCTGGTCAGCAGCGCCCAAGCAACGCGCGG-----TACACCCCG 8903
 Qy 302 GlyAspGlnGlyTrpGlyArgValThrLeuAsnLysSerLeuAsnValAlaTyrValAsn 321
 Db 8904 TACCAGGGGCGCGCGCGCGGTCTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8963
 Qy 322 GluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAlaGly 341
 Db 8964 ACCAGACCGCTACTCC-----GGCTTCCAC----- 8990
 Qy 342 LysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThr-----Thr 358
 Db 8991 -----ACGTGCG 9029
 Qy 359 AlaSerTyrThrLeuValAsnAsp-----LeuAspLeuValIle-----Thr 372
 Db 9030 GTGAGGTACACCAAGCTCGCGCAGCGCGCGCGGTACCTCAACCTGGCGGTCAACGGCAC 9089
 Qy 373 AlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnTrp 392
 Db 9090 GTCCCGCGCGGTG-----TTCAGCCTCTCCGAGGACCAT----- 9125
 Qy 393 AspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrIle 412
 Db 9126 -----GTACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9155
 Qy 413 IleGluVal---GlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
 Db 9156 GTCACCTCAGCGCGCTCTGGCAGGCTGGCGGCGCGCGCGCGCGCGCGCGCGCGCG 9215

```
Db 913 TATCAATCGGTCTCCACAGCTGCAAGCAAAAGTTATTACAGTTGGAGCGGTTGACAAAG 972
Qy 187 TyArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGlnPheSerSer 206
Db 973 TAT-----GATGTTATTAACAAGCTTCTCAAGC 999
Qy 207 ArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaProGlyThrPheIle 226
Db 1000 AGAGGCCCAACTGCACACGGCAGGCTTAAGCTGAGGTTGTTGCTCCAGGAACCTGGATA 1059
Qy 227 LeuSerAlaArgSerSerLeuAlaProAspSerPheThrAlaAsnTyrAsnSerLys 246
Db 1060 ATTGCTGCCAGCAAGT-----GGAACTAGCATGGGTCAACCAATTAATGACTAT 1110
Qy 247 TyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGln 266
Db 1111 TACACAGCAGCTCTCGGACATCAATGGCAACTCTCACTAGCTGGTATTGACGCCCTC 1170
Qy 267 LeuArgGluHisPheIleLysAsnArgGlyIleThrPro-----LysProSerLeu 283
Db 1171 TTGCTCAA-----GCACACCCGAGCTGGACTCCAGACAAAGTAAACACGCCCTC 1221
Qy 284 IleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGlyTyrProAsnGly 302
Db 1222 ATAGAAACTGCTGATATCGTAAAGCCAGATGAATAGCCGATATAGCCTAC----- 1272
Qy 303 AspGlnGlyTrpGlyArgValThrLeuAsnLysSerLeuAsnValAlaTyrValAsnGlu 322
Db 1273 -----GGTGCAGGTAGGTTAATGCATCAAGGCTATAAAC-----TACGATAACTAT 1320
Qy 323 AlaThrAlaLeuAlaThrGlyGlnLysAla-----ThrTyrSerPheGln 337
Db 1321 GCAAGCTAGTGTCTACTGGATATGTTGCCAACAAAGCGAGCCAAACTCACCAGTTGGT 1380
Qy 338 AlaGlnAlaGlyLysProLysIleSerLeuValThrThrAspAlaProGlySerThr 357
Db 1381 ATTAGCGGAGCTCGTTCGTAAGTGCACATTATCTGGGCAATGCCAAT----- 1431
Qy 358 ThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGln 377
Db 1432 -----ACGACCTTGATCTTTACTCTACGATCCCAATGGAAAC 1470
Qy 378 LysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsn 397
Db 1471 CAGGTT-----GACTACTCTTAC-----ACCGCTACTATGGA 1503
Qy 398 ValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrIleIleGluValGlnAla 417
Db 1504 TTCGAAAGGTTGGTTATTACACCCCACTGATGGACATGGACAATTAAAGTTGTAAAC 1563
Qy 418 TyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1564 TAC-----AGCGGAAGTGCAAACTATCAAGTAGATGTGTGTA 1599
```

RESULT 5

```
US-09-974-300-1934
; Sequence 1934, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 1934
; LENGTH: 1329
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1934
Alignment Scores:
Pred. No.: 9,3e-28 Length: 1329
Score: 341.00 Matches: 106
Percent Similarity: 44.58% Conservative: 42
Best Local Similarity: 31.93% Mismatches: 116
Query Match: 15.09% Indels: 68
DB: Gaps: 12
```

US-09-985-689A-4 (1-433) x US-09-974-300-1934 (1-1329)

```
Qy 11 AspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnGlnValValAlaValAlaAsp 30
Db 406 GAAGTGTGTGCAAAACAATCAGACGCTGACAGGCAAAAGGATGACAGTCGCTGTCATTGAT 465
Qy 31 ThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIle 50
Db 466 ACGGCGGTA-----TACCTCAGCAAGATCTTGAAGCAGATC 504
Qy 51 ThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaSerAspProAsnGlyHisGly 69
Db 505 AGGCTTTTCAAGACTTTATCAACAGAGAACAGAACCTATGATGACAAATGGCAGCGC 564
Qy 70 ThrHisValAlaGlySerValLeuGlyAsnAlaLeu-----AsnLysGly 84
Db 565 ACACACTGCGCGGTGATGCTTGCAGAACGAGGCGGCTCTCGGGTCAAGTACCGCGGA 624
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyLeuGly 104
Db 625 CCTGCTCTGAAGCAGAACTTGTGCGTGTAAAGTATTGGACAAAATGGGATCC---GGA 681
Qy 105 GlyLeu-----Pro 107
Db 682 TCGCTCGAAACCGTCATTCAGGCGTAGATTGGTGATTCATTCATTAAGAGAAATTCCT 741
Qy 108 SerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIleHisThrAsn 127
Db 742 GATGATCCGATCGACATTATTTCATGCTGCTGGTGCAGAGCCCTTGCCTACGAGAAT 801
Qy 128 SerTrpGlyAlaPro-----ValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
Db 802 GAAGAAGAAGATCCAGTCGTTAAAGCTGTTTCATGACGATGGGACGAGGC----- 852
Qy 144 ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly 163
Db 853 -----ATTGTTGATGTGGCGCAGCGGCAACTCCCGT 885
Qy 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 886 CCGTATGCGCAAAACGATTGCCAGCCGCGGTGTCAGACAGCAAGATATTATACAGTCGAGCC 945
Qy 184 ThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGln 203
Db 946 TTGATGACAG-----GATACAGTCAGCGGAGGATGACGATGTCGCTCT 993
Qy 204 PheSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaProGly 223
Db 994 TATTCAAGCAGAGGCGCGACAAATCTATGGTCAAGTCAAAACCGGACTTGTGTTACCGGCGC 1053
Qy 224 ThrPheIleLeuSerAlaArgSer-----SerLeuAlaProAspSerSerPheTrpAla 241
Db 1054 ACAATATTACGTCGCTTCGTTTCCCGGATCTTTTCTCGATAAGCTGCAAAAAACAAC 1113
Qy 242 AsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAla 261
Db 1114 AGAGTCGGCACAATAATATATGACATTGTCGGAACCTCGATGGTGTACGCCGATCTGGCA 1173
Qy 262 GlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysPro 281
```

Db 217 TCATACGACCTGGTACTGGAGCGCAAGTAATGGCAAGTACAGGAATGGCTCCAGCA 276
Qy 89 AlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGlyLeuGlyLeu 106
Db 277 GCTAAGCTGGCGGAATTAAGTTCTAGTGGCGGATGGTTCTGGAAGCATATCTACTATA 336
Qy 107 ProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIleHisThr 126
Db 337 ATTAAGGAGTGTGAGTGGCGGTGTGATAACAAGATAGTACGGAATTAAGTCATTAT 396
Qy 127 AsnSerTrpGlyAlaProValAsnGlyAlaTrpThrAlaAsnSerArgGlnValAspGlu 146
Db 397 CTTTCTCTGGTTCACGACGAGCTCAGATGGTACTCAGCTCTAAGTCAGCGCTGTTAAT 456
Qy 147 TyrValArgAsnAspMetThrValLeuPheAlaGlyAsnGluGlyProAsnSer 166
Db 457 GCACGGTGGGATGCTGATGATTGTTGTTGGTCCCGCTGGAACAGTGGACCTAACAG 516
Qy 167 GlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsn 186
Db 517 TATACAAATCGTTCTCCAGCAGCTGCAAGCAAGTTATTACAGTTGGAGCGGTGACAAG 576
Qy 187 TyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGlnPheSerSer 206
Db 577 TAT-----GATGTTATAACAAGCTTCTCAAGC 603
Qy 207 ArgGlyAlaThrArgAspGlyValIleLysProAspValThrAlaProGlyThrPheIle 226
Db 604 AGAGGCCCACTGCAGACGGCAGCTTAAGCTGAGGTTGTTGCTCCAGGAAGTGGATA 663
Qy 227 LeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnTyrAsnSerLys 246
Db 664 ATTGTCGCAGACGAAGT-----GGAAGTACGATGGTCAACCAATTAATGACTAT 714
Qy 247 TyrAlaTrpMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGln 266
Db 715 TACACAGCAGCTCTCGGACATCAATGGCACTCTCAGTAGCTGTATTGAGCCCTC 774
Qy 267 LeuArgGluHisPheIleLysAsnArgGlyIleThrPro-----LysProSerLeu 283
Db 775 TTGCTCCAA-----GCACACCGAGCTGGACTCCAGACAAAGTAAATAACAGCCCTC 825
Qy 284 IleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly--LeuGlyTyrProAsnGly 302
Db 826 ATAGAACTGCTGATATCGTAAAGCAGATGAATAGCGATATAGCTTAC-----876
Qy 303 AspGlnGlyTrpGlyArgValThrLeuAsnLysSerLeuAsnValAlaTrpValAsnGlu 322
Db 877 -----GGTCAGGTAGGTTAATGCATACAGGCTATAAC-----TACGATAACTAT 924
Qy 323 AlaThrAlaLeuAlaThrGlyGlnLysAla-----ThrTyrSerPheGln 337
Db 925 GCAAGCTAGTGTTCACATGGATATGTTGCCAACAAAGCGACCCAACTCACCAGTTCGTT 984
Qy 338 AlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThr 357
Db 985 ATTAGCGGAGCTTCGTTTCGTAACGCCACATTATACCTGGGCAATGCCAAT-----1035
Qy 358 ThrAlaSerTyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGln 377
Db 1036 -----AGGACCTTGATCTTTACCTCTACGATCCCAATGGAAAC 1074
Qy 378 LysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsn 397
Db 1075 CAGGTT-----GACTACTCTTAC-----ACCGCTACTATGGA 1107
Qy 398 ValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrIleIleGluValGlnAla 417
Db 1108 TTCGAAAAGGTTGGTTATTACACCCCACTGATGGAACATGGCAATTAAGTTGTAAGC 1167
Qy 418 TyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432

Db 1168 TAC-----AGCGGAAGTGCAAACTATCAAGTAGATGGTA 1203

RESULT 4

US-10-090-624-15
; Sequence 15, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:

; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15

LENGTH: 1962

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Synthetic

US-10-090-624-15

Alignment Scores:

Pred. No.:	6,82e-36	Length:	1962
Score:	416.50	Matches:	137
Percent Similarity:	44.14%	Conservative:	55
Best Local Similarity:	31.49%	Mismatches:	172
Query Match:	18.43%	Indels:	71
DB:	13	Gaps:	17

US-09-985-689A-4 (1-433) x US-10-090-624-15 (1-1962)

Qy 16 AsnTyrGlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 454 AACTTGGGATATGATGGTTCTGGAATCAATGGAATAATTGACACTGGAATTGAC---510
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeu-----53
Db 511 -----GCTTCTATCCCATCTCCAGGAAAAGTAATTTGGTGGGTAGAT 555
Qy 54 TyrAlaLeuGlyArgThrAsnAsnAlaSerAspProAsnGlyHisGlyThrHisValAla 73
Db 556 TTTGTCAATGGTAGG---AGTTATCCATACGATGACCATGACATGGACATCATGTAGCT 612
Qy 74 GlySerValLeuGly-----AsnAlaLeuAsnLysGlyMetAlaProGln 88
Db 613 TCAATAGCAGCTGGTACTGGAGCACAAGTAATGGCAAGTACAGGGAATGGCTCCAGA 672
Qy 89 AlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyLeuGlyLeu 106
Db 673 GCTAAGCTGGCGGAATTAAGGTTCTAGGTGCCGATGGTTCTTGGGAAGCATATCTACTATA 732
Qy 107 ProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIleHisThr 126
Db 733 ATTAAGGAGGATTGAGTGGGCGCGTGTATACAAAGATAAGTACGGAATTAAGGTCATTAT 792
Qy 127 AsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGlu 146
Db 793 CTTTCTCTTGGTTCAAGCCAGAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTTAAT 852
Qy 147 TyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlyGlyProAsnSer 166
Db 853 GCACGGTGGGATGCTCGGATTAGTTGTTGTTGGCTGGGAAACAGTGGACCTAACAG 912
Qy 167 GlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsn 186

US-09-985-689A-4 (1-433) x US-10-090-624-11 (1-1977)

Qy 8 ValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnGlyValValAla 27
Db 433 ATAGGGCGCGATACCGTCTGGAACTCCCTCGGTACGACGGAAGCGGTGTGGTGGC 492

Qy 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArg 47
Db 493 ATCGTCGATACGGGTATAGCGGAAC-----CACCCCGATCTGAAG 534

Qy 48 GlyLysIleThrAlaLeuTyr-----AlaLeuGlyArgThrAsnAsnAlaSerAspProAsn 66
Db 535 GGCAAGGTATAGGCTGGTACGACGCGCTCAACGCGAGGTGCGACCCCTACGATACCCAG 594

Qy 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
Db 595 GGCACGGAAACCCACGTTGCGGGTATCGTTGCCGGAACCGGACGCGTAACTCCACGATAC 654

Qy 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 100
Db 655 ATAGGGGTGCGCCCGCGCGAAGCTCGTCCGGTCAAGGTTCTCGGTCCCGACGCTTCG 714

Qy 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
Db 715 GGAAGCGTCTCCACCATCATCGCGGTGTGACTGGGTCTGTCAGAACAGGACAGATAC 774

Qy 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
Db 775 GGGTAAGGTCTATCAACTCTCCCTCGGTCTCTCCAGAGCTCCGACGGAACCGACTCC 834

Qy 141 SerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaGly 160
Db 835 CTGAGTCAGCGCTCAACAGCGCTGGAGCGCGGTATAGTCTGCTCGCGCGCGC 894

Qy 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 895 AACAGCGCGCGAACCCTACCGTCCGCTCACCGCGCGCGCGAAGGTCTATAACC 954

Qy 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
Db 955 GTCGGTGCA-----GTTGACAGCAACGACAAAC 981

Qy 201 IleAlaGlnPheSerArgGlyAlaThrArgAspGlyValGlyIleLysProAspValThr 220
Db 982 ATGCCAGCTTCTCCAGCAGGGACCGACCGCGGAGGCTCAAGCGGAGTCTGTC 1041

Qy 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerIleAlaProAspSerSerPheTrp 240
Db 1042 GCGCCCGCGCTTGACATCATAGCCCGCGCGCGCAGC-----GGAACCGAGCATGGCG 1092

Qy 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
Db 1093 ACCCGGTAACGACTACTACCAAGCGCTCTGGAACCGACGATGGCCCGCGCGAGTT 1152

Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro---- 279
Db 1153 TCGGGCGTGGCGCGCTCATCTCCAG-----GCCACCGCGAGCTGGACCCCGGAC 1203

Qy 280 -----LysProSerIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly--- 296
Db 1204 AAGGTGAAGACCGCCCTCATCGAGACCGCGCATAGTCGCCCGCCCAAGGAGATAGCGGAC 1263

Qy 297 LeuGlyTyrProAsnGlyAspGlnGlyTrpGlyArgValThrLeuAsnLysSerLeuAsn 316
Db 1264 ATCGCCTAC-----GGTGGGGTAGGTGAACGCTCTACAAGGC----- 1302

Qy 317 ValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGln-----Lys 331
Db 1303 ATCAAGTACGACGACTACGCAAGCTCCTTACCGGCTCTCCCGCGACAGGGAAGC 1362

Qy 332 AlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThr 351
Db 351

Db 1363 GCCACCCACACCTTCGACGTCAGCGGCGCCACCTTCGTGACCGCCACCTCTACTGGGAC 1422

Qy 352 AspAlaProGlySerThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIle 371
Db 1423 -----ACGGGCTCGAGCGACATCGACCTCTACCTC 1452

Qy 372 ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsn 391
Db 1453 TACGACCCCAAC-----GGGAACGAGGTTGACTACTCTCTACCGCGCTAC 1497

Qy 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyr 411
Db 1498 TAC-----GGCTTCGAGAGGTCGGTCTACTACACCGCGCGGACCTGG 1545

Qy 412 IleIleGluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
Db 1546 ACGETCAAGGTCGTCACTAC-----AAGGGCGCGCGAAGTACCAGGTCGACGTC 1596

Qy 432 Val 432
Db 1597 GTC 1599

RESULT 3

US-10-090-624-2
; Sequence 2, Application US/10090624
; Publication No. US2002013233A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozi
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-2

Alignment Scores:
Pred. No.: 3,47e-36 Length: 1236
Score: 416.50 Matches: 137
Percent Similarity: 44.14% Conservative: 55
Best Local Similarity: 31.49% Mismatches: 172
Query Match: 18.43% Indels: 71
Gaps: 17

US-09-985-689A-4 (1-433) x US-10-090-624-2 (1-1236)

Qy 16 AsnTyrGlyLeuTyrGlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThr 35
Db 58 AACTTGGGATATGATGGTCTGGAATCATATGAATATGACATCTGGAATTGAC--- 114

Qy 36 GlyArgAsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeu----- 53
Db 115 -----GCTTCTCATCCAGATCTCCAGGAAAAGTAATTGGTGGGTAGAT 159

Qy 54 TyrAlaLeuGlyArgThrAsnAsnAlaSerAspProAsnGlyHisGlyThrHisValAla 73
Db 160 TTTGTCAATGTAGG---AGTTATCCATACCATGACATGACATGGAACATCATGTAGCT 216

Qy 74 GlySerValLeuGly-----AsnAlaLeuAsnLysGlyMetAlaProGln 88

```

; PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 1305
; TYPE: DNA
; ORGANISM: Bacillus sp. KSM-KP43
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1305)
; OTHER INFORMATION:
US-10-385-662-1

Alignment Scores:
Pred. No.: 1 48e-210 Length: 1305
Score: 197.50 Matches: 379
Percent Similarity: 94.01% Conservative: 29
Best Local Similarity: 87.33% Mismatches: 25
Query Match: 87.94% Indels: 1
DB: 15 Gaps: 1

US-09-985-689a-4 (1-433) x US-10-385-662-1 (1-1305)
Qy 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnAsnAspTyrGlyLeuTyr 20
Db 1 AATGATGTTGCGGTGGAATTTGCAAGCGGATGTGCTCAGACGAGCTACCGGTTGTAT 60
Qy 21 GlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThrGlyArgAspSer 40
Db 61 GGACAGGACAGATCGTAGGGTTGCGATACAGGCTTGATACAGTCCCAATCAGAT 120
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCGATGATGAAGCTTCGCGGGAATTTACTGCAATATATGATGCGAGCGAGAT 180
Qy 61 AsnAlaSerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 181 AATGCCAATGATACGATGTCATGTCATGTCGCTGGCTCCGATTAGGAACGGC 240
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 241 TCCACTAATAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCAATCTATCATGGATAGC 300
Qy 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 301 GGTGGGGGACTTGGAGGACTACCTTCAATCTGCAACCTTATTCAGCCAGCATACAGT 360
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAla 139
Db 361 GCTGGTGCAGATTTACAACTCTCTGGGGAGCAGCAGTGAATGGGGCTTACACAACA 420
Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAla 159
Db 421 GATTCAGAAATGTGGATGATATGCGCAAAATGATGAGCATCTTTTCGTGTC 480
Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 481 GGGAAATGAAGCCGACCGCGGAACCATCAGTGCACCGAGCAGCTAAATGAATA 540
Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 541 ACATGCGAGATGACGAAACCTCGCCCAAGCTTTGGGCTTATGCGGCAATATCAAC 600
Qy 200 HisIleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 601 CATGTGGCACAGTTCTCTTCAGTGGACCGCAAAAGGATGACGAGTCAAAACCGGATGC 660
Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
Db 661 ATGCACCGGAAGCTTCATCTACTACGACAGATCTTCTCTTGCACCGGATTCCTCCTC 720
Qy 240 TrpAlaAsnTyrAsnSerLysThrAlaTyrMetGlyThrSerMetAlaThrProIle 259

```

```

; TGGCGGAACCATGACAGTAATAATGATATGATGGTGAACGCTCCATGGCTACACGATC 780
; ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
; GTTGTGGAACCGTGGCAGCAGCTTCGTGAGCATTTGTGAAACACAGAGGCATCACACCA 840
; LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
; AAGCCTTCTCTATTAAAGCGGCACTGATTCGCGTGCGAGCTGACATCGGCTTGGCTAC 900
; ProAsnGlyAspGlnGlyTyrGlyArgValThrLeuAsnLysSerLeuAsnValAlaTyr 319
; CCGAACGGAATCAACAGGATGGGACGAGTGACATTTGGATAAATCCCTGAACGTTGGCTAT 960
; ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
; GTGAACGAGTCCAGTTCTCTATCCACGAGCAAAAGCGAGCTACTCGTTTACTGTACT 1020
; AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrAla 359
; GCCGCAAGCTTTGAAAATCTCCCTGGTATGCTGTATGCCCTCGGAGCAACTGCT 1080
; SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
; TCCGTAACGCTTGTCAATGATCTGGACCTTGTCTATTACCGCTCCAAATGGCACACGAT 1140
; ValGlyAsnAspPheSerTyrProTyrAspAsnTrpAspGlyArgAsnValGlu 399
; GTAGGAATGACTTACTTCGCCATACATGATACTGGGATGGGCAATAACGTAGAA 1200
; AsnValPheIleAsnAlaProGlnSerGlyThrTyrIleIleGluValGlnAlaTyrAsn 419
; AATCTATTATTATGATGCCACCAAAAGCGGAGCTATACAAATTGAGGTACAGGCTTATAAC 1260
; ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
; GTACCGGTTGACACACAGACCTTCTCGTGGCAATTGTGAAT 1302

```

```

RESULT 2
US-10-090-624-11
; Sequence 11, Application US/10090624
; Publication No. US2002013233A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-11

Alignment Scores:
Pred. No.: 2 83e-38 Length: 1977
Score: 438.00 Matches: 136
Percent Similarity: 45.80% Conservative: 66
Best Local Similarity: 30.84% Mismatches: 171
Query Match: 19.38% Indels: 68
DB: 13 Gaps: 16

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2004, 08:10:21 ; Search time 327.248 Seconds
(without alignments)
4948.852 Million cell updates/sec

Title: US-09-985-689A-4

Perfect score: 2260

Sequence: 1 NDVARGIVKADVAQNYYGLY.....EVQAVNVPSGQRFSLAIVH 433

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2466186 seqs, 1870095128 residues

Total number of hits satisfying chosen parameters: 4932372

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

MODEL=frame+ p2n.model -DEV=xlh
Q-/cgn2_1/USPTO spool/US09985689/runat 31032004 161809 4271/app query.fasta_1.3498
-DB=PublishedApplications_NA -QWTF=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -PAIRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -MUSER=US0985689 @CGN 1 1 601 @runat 31032004 161809 4271
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PTC_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PTCUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description

RESULT 1
US-10-385-662-1
; Sequence 1, Application US/10385662
; Publication No. US20040002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO NORUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAEKI, KATSURISA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938U0
; CURRENT APPLICATION NUMBER: US/10385,662
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231

ALIGNMENTS

	1	1987.5	87.9	1305	15	US-10-385-662-1
2	438	19.4	1377	13	US-10-090-624-11	
3	416.5	18.4	1236	13	US-10-090-624-2	
4	416.5	18.4	1962	13	US-10-090-624-15	
5	341	15.1	1329	9	US-09-974-300-1934	
6	334.5	14.8	135638	14	US-10-314-657-1	
7	328	14.5	3624	14	US-10-156-761-5701	
8	328	14.5	9025608	14	US-10-156-761-1	
9	325	14.4	3417	14	US-10-156-761-1	
10	325	14.4	9025608	14	US-10-156-761-1	
11	313	13.8	4765	13	US-10-090-624-5	
12	300.5	13.3	3743	10	US-09-927-827-29	
13	298	13.2	1560	15	US-10-084-846A-113	
14	298	13.2	59816	15	US-10-084-846A-1	
15	298	13.2	59816	15	US-10-084-846A-2	
16	277.5	12.3	3303	14	US-10-156-761-5384	
17	277	12.3	3788	10	US-09-927-827-33	
18	276.5	12.2	2166	12	US-10-344-231-17	
19	276.5	12.2	2166	12	US-10-363-332A-17	
20	270	11.9	1140	8	US-08-322-678-11	
21	270	11.9	1140	16	US-10-323-324-11	
22	270	11.9	1143	14	US-10-313-853-6	
23	270	11.9	2588	12	US-10-344-231-20	
24	270	11.9	2588	12	US-10-363-332A-20	
25	269	11.9	1140	8	US-08-322-678-12	
26	269	11.9	1140	16	US-10-323-324-12	
27	263	11.6	840	14	US-10-203-812-1	
28	261.5	11.6	1306	9	US-09-966-921A-1	
29	261.5	11.6	1330	9	US-09-966-921A-5	
30	260.5	11.5	1485	9	US-09-974-300-1938	
31	259.5	11.5	2287	12	US-10-344-231-18	
32	259.5	11.5	2287	12	US-10-363-332A-18	
33	257.5	11.4	1971	9	US-09-974-300-1935	
34	248	11.0	3884	10	US-09-927-827-34	
35	247.5	11.0	3452	12	US-10-424-599-112429	
36	245.5	10.9	2192	10	US-09-927-827-30	
37	243	10.8	1140	15	US-10-146-905A-9	
38	239	10.6	1140	9	US-09-920-118-13	
39	232	10.3	1332	14	US-10-156-761-5689	
40	231	10.2	1575	12	US-10-344-231-19	
41	231	10.2	1575	12	US-10-363-332A-19	
42	228	10.1	1457	14	US-10-202-339-1	
43	227.5	10.1	522	10	US-09-824-893A-98	
44	227	10.0	1497	8	US-08-322-678-6	
45	227	10.0	1497	9	US-09-960-854B-1	

Sequence 1, Appli
Sequence 11, Appli
Sequence 2, Appli
Sequence 15, Appli
Sequence 1934, Ap
Sequence 1, Appli
Sequence 5701, Ap
Sequence 3306, Ap
Sequence 5, Appli
Sequence 29, Appli
Sequence 11, App
Sequence 1, Appli
Sequence 2, Appli
Sequence 5384, Ap
Sequence 33, Appli
Sequence 17, Appli
Sequence 11, Appli
Sequence 6, Appli
Sequence 20, Appli
Sequence 12, Appli
Sequence 12, Appli
Sequence 1, Appli
Sequence 5, Appli
Sequence 1938, Ap
Sequence 18, Appli
Sequence 18, Appli
Sequence 1935, Ap
Sequence 34, Appli
Sequence 30, Appli
Sequence 112429,
Sequence 9, Appli
Sequence 13, Appli
Sequence 5689, Ap
Sequence 19, Appli
Sequence 19, Appli
Sequence 1, Appli
Sequence 98, Appli
Sequence 6, Appli

Db 1328 GGCATGGACACACCGGAGACCGACCGCTGGAGGGGGG----- 1366
 Qy 144 ValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGlyAsnGly 163
 Db 1367 GTCGACAAAGCTGCCCGGAGAGGGCTCTCTGTCGCATCGCGCGGCAACGAGGGC 1426
 Qy 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
 Db 1427 CCGGAGTCG-----ATCGGTTCCCGCGGAGCGGACCGCGCTTCCCTCGCGGGC 1480
 Qy 184 ThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGln 203
 Db 1481 -----GTCGACGACAAAGCAAGCTCGCGGAC 1507
 Qy 204 PheSerSerArgGly---AlaThrArgAspGlyArgIleIleYsProAspValThrAlaPro 222
 Db 1508 TTCTCTCTCCACCGCGCCCGCCCTCGCGGAGCGGCCCATCAAGCGGACGTCACCGCTCCC 1567
 Qy 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
 Db 1568 GCGGTGGACATCACGGCGCGCTCGCGGAGGGCAACGACATCGCGGAGGTCGGTGAG 1627
 Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 1628 GGACCGCGCGGTACATGACCATCTCGGCGACGTCGATGCGGACCGCGACGTCGCGGGC 1687
 Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSer 282
 Db 1688 GCGCGCGCTCTGAGAGGAGCAG-----CACCCCGGAC 1720
 Qy 283 LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
 Db 1721 TGGACCTCCGCGCAACTGAAGGGCGCGCTCACCGGTCTCCACCAAGGGCGGC---AAGTAC 1777
 Qy 300 ProAsnGlyAspGlnGlyTyrGlyArgValThrLeuAsnLysSerLeuAsnValAlaTyr 319
 Db 1778 ACCCGGTCGAGCAGGTTTCGGCGCGGATTCAGCGCGACAGGGCGCTCCAGCAGACCGTG 1837
 Qy 320 ValAsnGluAlaThrAlaLeuAlaThrGly----- 329
 Db 1838 ATCGCGACCGCGTCTCGGTGAGTTCGGGTCTCAGCAGTGGCGGACACCGGACGACGAG 1897
 Qy 330 -----GlnLysAlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIle 346
 Db 1898 CCGGTACCAAGCAGCTGACCTACCGCAACCTCGGACCCAGGACGTCACGCTGAAGCTG 1957
 Qy 347 SerLeuValTThrAspAlaProGlySerThrThrAlaSer-----TyrThrLeuVal 364
 Db 1958 ACGTCACCGCCACCGACCCCAAGGCAAGCGCGCCCGCGGCTTCTTTCACGCTGGGC 2017
 Qy 365 AsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPhe 384
 Db 2018 -----GCCACACCGGTGACCTCCCGCGGGCGGC----- 2047
 Qy 385 SerTyrProTyrAspAsnAsnTyrAspGlyArg-----AsnAsnValGlu 399
 Db 2048 AGCGCTCCGTCGACATGACCGCGACACCGCGGTTCGGCGGACGCGTGGACGCGCGTAC 2107
 Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrIle----- 412
 Db 2108 TCGCGGTACGTGGTTCGCGGCGGGCGGCGAGACGCTCCGACGCGCGCGGTGCAG 2167
 Qy 413 IleGluValGlnAlaTyrAsnVal-ProSerGly 423
 Db 2168 GCGAGAGTTCGAGTCGACGAGTGCCTCCGCGC 2201

Search completed: April 4, 2004, 12:02:30
 Job time : 90.4512 secs

```
QY 204 PheSerSerArgGly---AlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
DB 1508 TTCTCTCCACGGCGCCCGCTCGGACGCGGCATCAGCCGAGTTCACGGCTCCC 1567
QY 223 GlyThrPheIleuSerAlaArgSerSerLeuAlaProAspSerPheTrpAlaAsn 242
DB 1568 GCGGTGGACATCACGGCCCTCGCGGAGGGCAACAGCATCGGCGCAGGAGTGGTGAG 1627
QY 243 TyrAsnSerLysTyAlaTrpMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
DB 1628 GGACGGCGCGGTATGACCATCTCCGGCACGTGATGGCACCCCGACCGTCCGGCGGC 1687
QY 263 AsnValAlaGlnLeuArgGlyHisPheIleLysAsnArgIleThrProLysProSer 282
DB 1688 GCGGGCGCCCTCTGAAGCAGCAG-----CACCCCGAC 1720
QY 283 LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
DB 1721 TGGACCTCGCGCAACTGAAGGCGCGCTCACCGCTCCACAGGGCGGC---AAGTAC 1777
QY 300 ProAsnGlyAspGlnGlyTrpGlyArgValThrLeuAsnLysSerLeuAsnValAlaTyr 319
DB 1778 ACCCGTTGAGCAGGTTTCGGCGGATCCAGGCGGACAGGCGCTCCAGCAGACCGTG 1837
QY 320 ValAsnGluAlaThrAlaLeuAlaThrGly----- 329
DB 1838 ATCGCGCAGCCCGTCTCGGTGAGCTTCGCGCTCCAGCAGTGGCCGACACCGACGACGAG 1897
QY 330 -----GlnLysAlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIle 346
DB 1898 CCGGTACACGACGAGTACCTACCGCACTCCGACCTCGGACCCAGGAGTCAAGCTG 1957
QY 347 SerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----TyrThrLeuVal 364
DB 1958 AGCTGACCGCCACCCAGCCGCGGCAAGGCGGCGCGCGGCTTCTTCACGTGGGC 2017
QY 365 AsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPhe 384
DB 2018 -----GCCACACGGTGACCGTCCCGGGGCGGC----- 402
QY 385 SerTyrProTyrAspAsnAsnTrpAspGlyArg-----AsnAsnValGlu 399
DB 2048 AGCGCTCGGTGCATGACCGCGCACACCGCGCTCGGCGGCGCACGCGTGGACGCGGTAC 2107
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrIle----- 412
DB 2108 TCGCGTACGTGGTCCCGCACCGGCGGCGGCGGACGTCGCGACGCGCGCGCGGTGCAG 2167
QY 413 IleGluValGlnAlaTyrAsnVal-ProSerGly 423
DB 2168 CGCGAGGTGAGTGTGACGACGTGACCGTCCGCGC 2201

RESULT 15
US-09-514-340-1
; Sequence 1, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: AKIRA ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
; ITS EXPRESSION PRODUCT
;
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
```

```
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICANT NUMBER: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN: <Unknown>
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-514-340-1
Alignment Scores:
Pred. No.: 3 41e-24 Length: 2809
Score: 340.50 Matches: 131
Percent Similarity: 40.93% Conservative: 54
Best Local Similarity: 28.98% Mismatches: 171
Query Match: 15.07% Indels: 96
DB: 4 Gaps: 18
US-09-985-689A-4 (1-433) x US-09-514-340-1 (1-2809)
QY 18 GlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
DB 998 GGCTACGACGCAAGGCGCGTGAAGATCGCGTCTCGACACCGGTGTGACACGAGC--- 1054
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
DB 1055 -----CATCGGACCTGAAGGCGGCGGTGACCGGTCCAGAACTTCACC 1099
QY 58 ArgThrAsnAsnAlaSerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
DB 1100 GCCCGCGCGCGCGCGCGCAAGGTGGCGCACGCCACCCACCGCTCGATCGCGCGCG 1159
QY 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
DB 1160 GGCACGGCGCCCGCATCCCAAGGGCAAGTACAGGGGTGCACCGCGCGCGCGATCCTC 1219
QY 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThr 112
DB 1220 AACGGCAAGGTCTCTCGACGACTCC-----GGTTTCGCGCAGCACTCCGGCATC 1267
QY 113 LeuPheSerGlnAlaTrpAsnAla-----GlyValaArgIleHisThrAsnSerTrpGly 130
DB 1268 CTCGCCGCGCATGGAGTGGCGCGCGCGCGACGCGCGCGCGCATGCTCAACATGAGCTCGGC 1327
QY 131 Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
```

```

Db 1595 GAGGGCAACGACATCGCGCAGGAGTCTGGTGAGGGACCGCGGCTACATGACCATCTCC 1654
Qy 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
Db 1655 GGCACGTCTGATGGCCACCCCGACGTCGCGGGCGGCGGCTCTCTGAGCAGCAG--- 1711
Qy 273 IleLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
Db 1712 -----CACGCCGACTCGACCTCCGCCGAACCTGAAGGGCGCG 1747
Qy 290 IleAlaGlyAlaThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTyrGlyArg 309
Db 1748 CTCACCGGCTCCACCAAG---GGCGGCAAGTACACCCCGTTGAGCAGGGTTCGGGCCCG 1804
Qy 310 ValThrLeuAspLysSerLeuAsnValAlaPheValAsnGlnThrSerSerLeuSer--- 328
Db 1805 ATCCAGGGCGACAAAGGCGTCCAGCAGACCGGTGATCGCGCACCGGCTCTCGGTGAGCTC 1864
Qy 329 -----ThrAsnGlnLysAlaThrTyrSer 336
Db 1865 GCGGTCCAGAGTGGCCGACACCGACGACGCGGTACCAAGCAGCTGACCTACCGC 1924
Qy 337 PheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValTyrSerAsp----- 353
Db 1925 AACCTCGGCACCCAGGACGTACGCTGAGCTGACGTGACCGCGCCACCGACCCCAAGGGC 1984
Qy 354 -----AlaProAlaSer-----ThrSerAlaSerValThrLeuValAsnLeuAsp 369
Db 1985 AAGGGCGGCCCGCGGGCTTCTTCACGCTGGCGGCCACCCAG----- 2026
Qy 370 LeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyr 389
Db 2027 -----GTGACCGTCCCGCGGCGCGCAGC---GCCTCCGTGACATGACCGGC----- 2071
Qy 390 AspAsnAsnTyrAspGlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGln 408
Db 2072 GACACCGGCTCGCGGCACCGGTGACGCGCGGTACTCGCGGTACGTGCTGCC---ACG 2128
Qy 409 SerGlyThrTyrThrVal-----GluValGlnAlaTyrAsn 420
Db 2129 GCGCGCGGCAGACGGTCCGACGCGCGCGGTGACGCGCGAGGTGAGTCGTACGAC 2188
Qy 421 Val-ProGlnGlyProGlnAla 427
Db 2189 GTGACCGTCCGCGCACATCGGCC 2210

```

Search completed: April 4, 2004, 01:12:51
 Job time : 364.684 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:50:44 ; Search time 63.5977 Seconds
(without alignments)
3787.066 Million cell updates/sec

Title: US-09-985-689A-6
Perfect score: 2250
Sequence: 1 NDVARGIVKADVAGSSVGLY.....EVQAYNVQGPQAFSLAIVN 434

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q/cgn2_1/USFT0_spool/US0995689/runat_31032004_161807_4184/app_query.fasta_1.3498
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORSE=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0995689 @CGN 1 1 142 @runat_31032004_161807_4184 -NCPUS=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A COMB.seq:
2: /cgn2_6/ptodata/2/ina/5B COMB.seq:
3: /cgn2_6/ptodata/2/ina/5A COMB.seq:
4: /cgn2_6/ptodata/2/ina/5B COMB.seq:
5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq:
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2143	95.2	1923	4	US-09-509-814A-5
2	2141	95.2	1923	4	US-09-509-814A-7
3	2122	94.3	1920	4	US-09-509-814A-3
4	2116.5	94.1	3003	2	US-08-873-479-41
5	450.5	20.0	1977	3	US-08-894-818B-2
6	450.5	20.0	1977	4	US-08-894-818B-2
7	406.5	18.1	1236	4	US-09-445-472-11
8	406.5	18.1	1236	4	US-09-445-472-2
9	406.5	18.1	1566	3	US-08-894-818B-4
10	406.5	18.1	1962	3	US-08-894-818B-34
11	405.5	18.0	1977	3	US-09-445-472-15
12	360.5	16.0	2539	3	US-08-894-818B-6
					US-09-000-016-3

13	360.5	16.0	2539	4	US-09-514-340-3	Sequence 3, Appli
14	355.5	15.8	2809	3	US-09-000-016-1	Sequence 1, Appli
15	355.5	15.8	2809	4	US-09-514-340-1	Sequence 1, Appli
16	318.5	14.2	2532	1	US-07-671-376C-4	Sequence 4, Appli
17	317	14.1	2835	1	US-08-750-532-2	Sequence 2, Appli
18	317	14.1	4765	1	US-08-750-532-8	Sequence 8, Appli
19	317	14.1	4765	3	US-08-894-818B-7	Sequence 7, Appli
20	317	14.1	4765	4	US-09-445-472-5	Sequence 5, Appli
21	305.5	13.6	1859	3	US-08-894-818B-15	Sequence 15, Appli
22	290	12.9	898	1	US-08-750-532-7	Sequence 7, Appli
23	286.5	12.7	564	1	US-08-750-532-11	Sequence 11, Appli
24	286.5	12.7	564	3	US-08-894-818B-14	Sequence 14, Appli
25	272	12.1	1306	4	US-09-966-921A-1	Sequence 1, Appli
26	272	12.1	1330	4	US-09-966-921A-5	Sequence 5, Appli
27	267	11.9	807	1	US-07-706-691G-56	Sequence 56, Appli
28	267	11.9	807	1	US-08-254-021-56	Sequence 56, Appli
29	267	11.9	807	2	US-08-618-446-56	Sequence 56, Appli
30	267	11.9	807	3	US-08-980-135-56	Sequence 56, Appli
31	267	11.9	807	4	US-09-585-798-56	Sequence 56, Appli
32	266.5	11.8	10216	2	US-08-875-154-1	Sequence 1, Appli
33	266	11.8	807	1	US-08-566-369-2	Sequence 2, Appli
34	266	11.8	807	3	US-09-074-331-2	Sequence 2, Appli
35	266	11.8	807	5	PCT-US95-01937-2	Sequence 2, Appli
36	265	11.8	807	1	US-07-706-691G-57	Sequence 57, Appli
37	265	11.8	807	1	US-08-254-021-57	Sequence 57, Appli
38	265	11.8	807	2	US-08-618-446-57	Sequence 57, Appli
39	265	11.8	807	3	US-08-980-135-57	Sequence 57, Appli
40	265	11.8	807	4	US-09-585-798-57	Sequence 57, Appli
41	264	11.7	840	1	US-08-434-255-7	Sequence 7, Appli
42	264	11.7	840	1	US-08-459-967-7	Sequence 7, Appli
43	264	11.7	840	1	US-08-460-327-7	Sequence 7, Appli
44	264	11.7	840	1	US-08-459-871-7	Sequence 7, Appli
45	264	11.7	840	3	US-09-024-532-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-509-814A-5
; Sequence 5, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUM
; APPLICANT: NOMURA, NASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 5
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-5

Alignment Scores: 1.53e-213 Length: 1923
Pred. No.: 2143.00 Matches: 406
Score: 2143.00
Percent Similarity: 97.93% Conservative: 19

```
Best Local Similarity: 93.55%
Query Match: 95.24%
DB: 4
Gaps: 0
Indels: 0
Mismatches: 9
```

US-09-985-689A-6 (1-434) X US-09-509-814A-5 (1-1923)

Qy	1	AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr	20
Db	619	AATGATGTTCCGGTGGAAATTGTCAAGCGGATGTGGCTCAGAGGAGCTACGGTTGTAT	678
Qy	21	GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer	40
Db	679	GGACAAGGACAGATCGTAGCGGTGCGGATACAGGGCTTGATACAGGTCCGAATGACAGT	738
Qy	41	SerMetHisGlnAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsn	60
Db	739	TCGATGATGAAGCTTCCCGGGAATTTACTGCAATTATATGCTATATGCGACGGACGAT	798
Qy	61	AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly	80
Db	799	AATGCCAATGATACGAATGTCATGTTAGCATGCTGGCTGGCTCCGTATTTAGGAACGGC	858
Qy	81	ThrSerAsnIysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSer	100
Db	859	TCCACTAATAAGGAATGGCGCTCAGCGAATCTAGTCTTCCAACTCATATGATAGAC	918
Qy	101	AsnGlyGlyLeuGlyGlyLeuProSerAsnValSerThrLeuPheSerGlnIleTyrSer	120
Db	919	GGTGGGGACITGGAGGACTACCTTCGAATCTTCMAACCTTATTACGCCAAGCATACAGT	978
Qy	121	AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrThr	140
Db	979	GCTGGTGCCAGAAATTCACAAACTCTCTGGGAGCAGCAGTGAATGGGCTTACACACA	1038
Qy	141	AspSerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAlaAla	160
Db	1039	GATTCAGAAATGTGGTAGCTATGTGGCAGAAAATGATATGACGATCTCTTCGTGTGC	1098
Qy	161	GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle	180
Db	1099	GGGAATGAAGACCCGAACGGCGGAACCATCAGTGCACCGAGCAGCTAAAAATGCAATA	1158
Qy	181	ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn	200
Db	1159	ACAGTCGGAGCTACGGAAAACTCCGCCCAAGCTTTGGGTCTTATGCGACAATATCAAC	1218
Qy	201	HisValAlaGlnPheSerSerArgGlyProThrIysAspGlyArgIleLysProAspVal	220
Db	1219	CATGTGGCACAGTTCTCTTCACGTGGACCGCAAAAGGATGACGATCAAAACCGGATGC	1278
Qy	221	MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe	240
Db	1279	ATGCAACCGGAAAGCTTCATACTATCAGCAAGATCTCTCTTGACCCGGATTCTCTCTTC	1338
Qy	241	TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle	260
Db	1339	TGGCGGAACCATGACAGTAAATATGTCATACATGGGTGGAACGTCCATGGGTACACGATC	1398
Qy	261	ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro	280
Db	1399	GTGTGTGAACCGTGGCCACAGCTTCGTGAGCATTTGTGAAAAACAGAGCATCACACCA	1458
Qy	281	LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaThrAspIleGlyLeuGlyTyr	300
Db	1459	AAGCCTTCTCTATTTAAAGCGCACTGATTCGGGTGCAGCTGACATCGGCCCTTGGGTAC	1518
Qy	301	ProSerGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe	320
Db	1519	CCGAACGGTAACCAAGGATGGGACCATGTACATTTGGATTAATCCCTGACAGTGGCTAT	1578
Qy	321	ValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAlaGln	340
Db	1579	GTGAACAGATCCAGTTCTCTATCCACCAAGCAAAAGCGAGCTACTCGTTTACTGTACT	1638

RESULT 2

```

US-09-509-814A-7
; Sequence 7, Application US/09509814A
; Patent No. 6376227
;
; GENERAL INFORMATION:
;
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUOSHI
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUM
; APPLICANT: NOMURA, MASAFUMI
;
; TITLE OF INVENTION: ALKALINE PROTEASE
;
; FILE REFERENCE: 0327-0832-0PCT
;
; CURRENT APPLICATION NUMBER: US/09/509,814A
;
; CURRENT FILING DATE: 2000-04-06
;
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
;
; PRIOR FILING DATE: 1998-10-07
;
; PRIOR APPLICATION NUMBER: JP 9-274570
;
; PRIOR FILING DATE: 1997-06-08
;
; NUMBER OF SEQ ID NOS: 24
;
; SOFTWARE: PatentIn version 3.0
;
; SEQ ID NO 7
;
; LENGTH: 1923
;
; TYPE: DNA
;
; ORGANISM: Bacillus sp.
;
; FEATURE:
;
; NAME/KEY: CDS
;
; LOCATION: {1}..(1923)
;
; US-09-509-814A-7

```

Alignment Scores:

Alignment scores:			
Pred. No.:	2,488-213	Length:	19233
Score:	2141.00	Matches:	406
Percent Similarity:	97.9%	Conservative:	19
Best Local Similarity:	93.55%	Mismatches:	9
Query Match:	95.1%	Indels:	0
DP:	4	Gaps:	0

US-09-985-689A-6 (1-434) x US-09-509-814A-7 (1-1923)

[illegible]

Db 739 TCGATGCATGAAGCCTTCGGCGGGAATAATTACTGCATTATATGCAATTCGGACGACGAAT 798
Qy 61 AsnAlaAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 799 AATGCCAATGATACGAATGGTCATGGTACGCAATGGCTGGCTCGTATATAGGAACGGC 858
Qy 81 ThrSerAsnLysGlyMetAlaProGlnAlaLeuValPheGlnSerValMetAspSer 100
Db 859 TCCACTAATAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCATCTCATGATGATAGC 918
Qy 101 AsnGlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyThr 120
Db 919 GGTGGGGACTTGGAGGACTACTTCGAACTCTGCAAACTTATTCAGCAAGCATACAGT 978
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrThr 140
Db 979 GCTGTGTCAGAAATTCATCAAACTCTCTGGGAGCAGCAGTGAATGGGGCTTACACACA 1038
Qy 141 AspSerArgAsnValAspAspTyThrValArgLysAsnAspMetAlaValLeuPheAla 160
Db 1039 GATTCAGAAATGTGATGACTATGTGCGCAAAATGATATGACGATCTTTTCGCTGCC 1098
Qy 161 GlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1099 GCGAATGAAGGACCGAAGCGGGAACCATCAGTCACCGAGGCACAGCTAAAAATGCAATA 1158
Qy 181 ThrValGlyAlaThrGlnAsnLeuArgProSerPheGlySerTyThrAlaAspAsnIleAsn 200
Db 1159 ACAGTCGGAGCTACGGAAACCTCGCCCAAGCTTTGGGTCTTATGCGGACAAATCAAC 1218
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1219 CATGTGGCAGCTTCCTTCACGTGGACCGCACAAAGGATGGACGATCAAAACCGGATGTC 1278
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1279 ATGGCACCGGACGCTTACTATATCAGCAAGATCTTCTTCGACCGGATTCCTCCCTTC 1338
Qy 241 TrpAlaAsnHisAspSerLysTyThrAlaTyMetGlyGlyThrSerMetAlaThrProIle 260
Db 1339 TGGCGCAACCATGACAGTAATAATGATACATGGGTGGGAACCTCCATGGCTACACCGATC 1398
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 280
Db 1399 GTTGCTGGAAACGTGGCAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCATCAACCA 1458
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaThrAspIleGlyLeuGlyTyThr 300
Db 1459 AACCTTCTCTATTAAGCGGCACCTGATTCGCGTGCAGCTGACATCGGCCCTTGGCTAC 1518
Qy 301 ProSerGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 320
Db 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATTGGATAAATCCCTGAACGTTGCCTAT 1578
Qy 321 ValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThrTySerPheThrAlaGln 340
Db 1579 GTGAACGAGTCCAGTTCTCTATCCACGACCAAAAGCGACGCTACTCGTTTACTGCTACT 1638
Qy 341 SerGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrSerAla 360
Db 1639 GCGCGCAAGCCTTTGAAATCTCCCTGGTATGCTGATGCTCCCTCGGACGACCACTGCT 1698
Qy 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrLysTyThr 380
Db 1699 TCGGTAAACGTTCTGATGATCTGGACCTTGTCAATACCGCTCCAAATGGCACACAGTAT 1758
Qy 381 ValGlyAsnAspPheThrAlaProTyAspAsnAsnTrpAspGlyArgAsnAsnValGlu 400
Db 1759 GTAGGAAATGACCTTACTTCGCCATACATGATAACTGGGATGGCGCGCAATACGTAGAA 1818
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyThrValGluValGlnAlaTyThrAsn 420
Db 1819 AATGTATTTATTAATGCACCAAGCGGACGCTATACAAATGAGTACAGGCTTATAAC 1878

Qy 421 ValProGlnGlyProGlnAlaPheSerLeuAlaIleValAsn 434
Db 1879 GTACCGTTGGACCACAGAACTTCTCGTTGGCAATGTGAAT 1920

RESULT 3

US-09-509-814A-3
; Sequence 3, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAKI, KATSUHISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUMI
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 3
; LENGTH: 1920
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1920)
US-09-509-814A-3

Alignment Scores:
Pred. No.: 2,37e-211 Length: 1920
Score: 2122.00 Matches: 400
Percent Similarity: 97.70% Conservative: 24
Best Local Similarity: 92.17% Mismatches: 10
Query Match: 94.31% Indels: 0
DB: 4 Gaps: 0

US-09-985-689A-6 (1-434) x US-09-509-814A-3 (1-1920)

Qy 1 AsnAspValAlaAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyGlyLeuTyThr 20
Db 616 AATGATGTGGCCAGAGATATTGTCAAAGCGGATGTGGCACAGACGAGCTACGGTTTGAT 675
Qy 21 GlyGlnGlyGlnValValAlaAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGACAAGGCCAGATGTGCGCAGTTGCCGATACTGGATTGGATACAGGAAGAAACGACGT 735
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaIleTyThrAlaLeuGlyAlaGlyThrAsn 60
Db 736 TCGATGATGAAGCCTTCCGCGGTAATAATACACACTATATGCACTGGGTGGACGAAT 795
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 796 AATCGAATGATACGAACGGTCACTGTCATGCCATGTGGCAGTTCGGTATTAGGAAATGGC 855
Qy 81 ThrSerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSer 100
Db 856 GCAACGAATAAAGGAATGGCACCTCAACCGCAATCTGGTTTTCATCATCATGATGATGC 915
Qy 101 AsnGlyGlyLeuGlyGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyThr 120
Db 916 AGTGGTGGCTTGGAGGCTTGCCTTCCCAATCTGCACAACTTATTTCACCAAGCATTCAGT 975
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThr 140

Db 976 GCAGTGGCCAGATTTCATACAACTCCTGGGGGAGCGGTGAATGGGGCTACAGACA 1035
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnMetAlaValLeuPheAla 160
Db 1036 GATTCAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1095
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 180
Db 1096 GGAATGAAGCGGACCGGCTACATCAGTGCACCTGGTACGGCTAAACCGCCATA 1155
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnLeu 200
Db 1156 ACAGTGGGGGACACCGGAAACCTGCTCAAGCTTCGTTCTCTATGACATATATTAAC 1215
QY 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1216 CACGTTGCACAGTTCCTCCCGTGGCCGACAAAGATGGGGGATCAAGCTGATGTC 1275
QY 221 MetAlaProGlyThrPheLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1276 ATGGCGCCAGGACATACATATTTATCAGCAAGATCTTCTTGCAACCGGATTCCTCTTC 1335
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
Db 1336 TGGCGGATCATGACAGCAATATGCTATATGGTGGGACGTCCATGGCAACCGATT 1395
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 280
Db 1396 GTTGGGGGATGTGACAGCTCCGAGCATTTTGTGAAAATAGAGGAATCACTCT 1455
QY 281 LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaThrAspIleGlyLeuGlyTyr 300
Db 1456 AAGCTTCCCTATTGAAAGCAGCTTGAATTCAGGTGCTGCTGATTTGGATTGGGTTAT 1515
QY 301 ProSerGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 320
Db 1516 CCGAACGAAACCAAGGATGGGCGGAGTGCACCTCGATAAATCGTTGAACGTTCCTAT 1575
QY 321 ValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAlaGln 340
Db 1576 GTGAACGAATCCAGTGCCTTATCACTAGCAAGCAAGCAATATACCTTTACTGCAACG 1635
QY 341 SerGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrSerAla 360
Db 1636 CGGGGCAAGCAATGAAATCTCCCTGATGATGGTGGATGCCCTGCAAGCACTCTGCT 1695
QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyr 380
Db 1696 TCTGTAACTCCCTGTCATGATTTGGATTGGTTCATTACAGCACCAACCGAAACAGATAT 1755
QY 381 ValGlyAsnAspPheThrAlaProTyrAspAsnTrpAspGlyArgAsnAsnValGlu 400
Db 1756 GTCCGGGATGATCTTCAGCAACCAATTTGACAACTACCTGGATGGCCCAATACGTAGAA 1815
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 420
Db 1816 AATGTATTTATTAATTCGCCCAAGTGGAAACATATACCATGAGTGGCAAGCATATAT 1875
QY 421 ValProGlnGlyProGlnAlaPheSerLeuAlaIleValAsn 434
Db 1876 GTCCGGTGGACCAACAACTTCTCGTTGGCAATTTGTAAC 1917

RESULT 4

US-08-873-479-41
; Sequence 41, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; APPLICANT: Lynne, Christanson
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:

ADDRESS: No. 5891701o No. 5891701disk of No. 5891701th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/873,479
FILING DATE: 12-JUN-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Agriis, Cheryl H
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 5251.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 3003 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-873-479-41

Alignment Scores:
Pred. No.: 1,81e-210 Length: 3003
Score: 2116.50 Matches: 400
Percent Similarity: 97.00% Conservative: 21
Best Local Similarity: 92.17% Mismatches: 12
Query Match: 94.07% Indels: 1
DB: 2 Gaps: 1

US-09-985-689A-6 (1-434) x US-08-873-479-41 (1-3003)

QY 1 AsnAspValAlaAaGGIyleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCCCGTGGCATTGTGAAGACAGACGTCGCAAAATAAATCTTGGCTTAT 1529
QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GGACAAGACAGATGTAGCAGTTGCTGATACTGGGCTTGATACGGAAGAAATGACAGT 1589
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCGATGCATGAAGCATTCGCGGTAAGATTACCCGCACTATATGCACTGGGCGAAGCAAT 1649
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 1650 AACGCCAATGATCCAAATGGACATGGAAACCCATGTTGCTGGATCTGTGTAGGAATGCT 1709
QY 81 ThrSerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSer 100
Db 1710 ACA---AATAAGGGATGGCCCGCAAGCCAACTCTAGTCTTTCAATCTATTATGATAGT 1766
QY 101 AsnGlyGlyLeuGlyGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSer 120
Db 1767 GGTGGAGGGCTGGGAGGACTACCTCTAATCTACAAACATTTATTCAGTCAAGCATATAGT 1826
QY 121 AlaGlyAlaAaGIIeHisThrAsnSerTrpGlyAlaProValAsnGlyValaTyrThrThr 140
Db 1827 GCTGAGCGAGANTTCACGAAATTCATGGGGGCTCCAGTAAACGGTGCCTATACGACA 1886
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAla 160
Db 1887 GACTCTCGAAATGTTGATGATATGATGATGATGATGATGATGATGATGATGATGATGAT 1946

161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaIleValAla 180
1947 GGAATGAGGACCGAGTACGATCAATCACTGCACAGGACGACCAAAATCGATT 2006
181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerThrAlaAspAsnIleAsn 200
2007 ACAGTTGGGGCAACCGAAACCTACGTCGATCGGATCTTATCGGATAATTAAC 2066
201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyValGlyIleLysProAspVal 220
2067 CATGTTGCTCAATCTCTTCACGAGTCTCTATAGATGGACGATTAAGCCGACGCTC 2126
221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
2127 ATGGCACCAGGACGATATCTCTCTGTAGATCATATAGCTCCAGATTCCTCATTC 2186
241 TrpAlaAsnHisAspSerLysThrAlaValMetGlyGlyThrSerMetAlaThrProIle 260
2187 TGGCAACCATGATAGTAATAGCTACATGCTGATGGTGGTACTTATGGCTACTCAATT 2246
261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 280
2247 GTAGCAGGTAATGTTCCACAAATTAAGGGAGCATTTGTGAAAATAGAGGGTAATCTCT 2306
281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaThrAspIleGlyLeuGlyTyr 300
2307 AAGCCTTCCTTTTAAAGCTGCTTTAATGCGAGTGTGCGGATGTTGACCTTGGCTTT 2366
301 ProSerGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 320
2367 CCAATGGTAACCAAGATGGGAGAGTAACGTTAGATAAATCCCTAATGTCGCAATT 2426
321 ValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThrThrSerPheThrAlaGln 340
2427 GTGAATGAACAGGAGCCCTTTATCAACAGTCAAAAAGCAACATATTCGTTTACGCTCAA 2486
341 SerGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrSerAla 360
2487 GCTGGTAACCTTTAAATATCACTGTTTGTGACAGATGCACCAAGTAGCAGCGCA 2546
361 SerValThrIleuValAsnAspLeuAlaValIleThrAlaProAsnGlyThrLysTyr 380
2547 TCATAACTTTAGTAATGATTTAGACTTAGTAATCACTGCACCAAAATGGAATCAATAC 2606
381 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 400
2607 GTCGGAATGACTTTACAGCACCCTATGATAACATTTGGATGGCAGAAACACGTTGGA 2666
401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 420
2667 AATGTTTATCAATGCTCTCAAGCGGAACGTATACAGTCGAAGTCGAGGCTTACAAT 2726
421 ValProGlnGlyProGlnAlaPheSerLeuAlaIleValAsn 434
2727 GTACCAAGTAGTCGCAACCTTTCTTCTAGCGATTGTACAT 2768

RESULT 5

US-08-894-818B-2
; Sequence 2, Application US/088948188
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyoza
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03953
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1977 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-2

Alignment Scores:
Pred. No.: 4,81e-37 Length: 1977
Score: 450.50 Matches: 138
Percent Similarity: 47.05% Conservative: 69
Best Local Similarity: 31.36% Mismatches: 170
Query Match: 20.02% Indels: 63
DB: 3 Gaps: 15

US-09-985-689A-6 (1-434) x US-08-894-818B-2 (1-1977)

QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnValValAla 27
DB 433 ATAGGGCCGATACCGTCTGGAACCTCCTCGGTACACGAGCGGTGTGTGTTGCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
DB 493 ATCTCGATACGGGTATAGACGCAAC-----CACCCCGATCTCTGAAG 534
QY 48 GlyLysIleThrAlaIleTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
DB 535 GCGAAGGTCTAGCTGTGTAGACCGCTCAACGCGAGGTGCGCCCTACGATGACGAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThrSerAsnLys----- 84
DB 595 GGCACAGCAACCCACGTTGCGGTATCGTTCGCGAACCAGCGCGTTAACTCCAGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMet-----AspSerAsn 101
DB 655 ATAGCGGTGCGCCCGCGGAGCTGTGCGCGTCAAGGTTCTCGGTGCGGACGTTGCG 714
QY 102 GlyGlyLeuGlyGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSerAla 121
DB 715 GGAAGCGTCTCCACCATCATCGCGGTGTTGACTGGGTGCGTCCAGAACAGGACAGTAC 774
QY 122 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAsp 141
DB 775 GGGATAAGGTGTCATCAACCTCTCCTCGGTCTCTCCAGAGCTCCGACGGAACCGACTCC 834

```

; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151989/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-11

Alignment Scores:
Pred. No.: 4,81e-37 Length: 1977
Score: 450.50 Matches: 138
Percent Similarity: 47.05% Conservative: 69
Best Local Similarity: 31.36% Mismatches: 170
Query Match: 20.02% Indels: 63
DB: Gaps: 15

US-09-985-689A-6 (1-434) x US-09-445-472-11 (1-1977)
QY 8 VallysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnValValala 27
Db 433 ATAGGGCGGATACCGTCTCGAAGCTCTCCGCTACGACGAGCGGTGTGTGGTGGC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATCTGTCGATACGGGTATAGACGCGAAC-----CACCCGATCTGAAG 534
QY 48 GlyIleThrAlaIleTyr---AlaLeuGlyArgThrAsnAlaAsnAspProAsn 66
Db 535 GCGAAGGTCATAGGTGTCGACGCGCTCAACGGCAGTTCGACCCCTCAGTACGACG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThrSerAsnLys----- 84
Db 595 GACACGCGGACCCACGTTGCGGGTATCGTTCGCGAACCGCGGTTAACTCCAGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMet-----AspSerAsn 101
Db 655 ATAGCGGTGCGCCCGCGCGAAGCTCTCGCGGTCAAGGTTCTCGGTGCGCGACGTTCC 714
QY 102 GlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSerAla 121
Db 715 GGAAGCGTCTCCACCATCATCGCGGTGTGTGCTGGTCCGACGACGACCAAGTAC 774
QY 122 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThrAsp 141
Db 775 GGGATAAGGGTCATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGACGCGAACCGACTCC 834
QY 142 SerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAlaGly 161
Db 835 CTCAGTCAGCGCGTCAACACCGCTGGGACGCGGTATAGTAGTCTCGGTCCGCGCGC 894
QY 162 AsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 181
Db 895 AACAGCGCGGCGAACAACCTACACCGTCGGCTCACCCTCGCGCGCGCGGACGAGTCA 954
QY 182 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 201
Db 955 GTCGGTGCA-----GTTGACAGCAACGACAAC 981
QY 202 ValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspValMet 221
Db 982 ATCCGACGCTTCTCCAGCAGGAGGACCGACCGCGGACGGAAGGCTCAAGCCGAGTCT 1041
QY 222 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPheTrp 241
Db 1042 GCCCGCGCGGTGACATCATAGCCCGCGCGCGAGC-----GGAACCAAGTGGGC 1092
QY 242 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 261
Db 1093 ACCCGGATAACGACTACTACCAAGGCTCTGGAAACCGACGATGCGCACCCGCGCGTT 1152
QY 262 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro--- 280
Db 1153 TCGGGCGTGGCGGCTCATCTCTCCAG-----GCCACCGAGCTGGACCCCGGAC 1203
QY 281 -----LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaThrAspIleGly--- 297
Db 1204 AAGGTGAAGACCGCCTCATCGAGACCGCGACATAGTCGCCCCCAAGGAGATAGCGGAC 1263
QY 298 LeuGlyTyrProSerGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeu--- 316
Db 1264 ATCGCCTAC-----GTTGCGGTAGGTGAACGTCTACAGGCGCATCAAG 1308
QY 317 -----AsnValAlaPheValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThr 334
Db 1309 TACGACGACTACGCGCAAGCTCACCTTCACCGGCTCCGTCGCGCGACAAGGAGCGCAAC 1368
QY 335 TyrSerPheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValTyrSerAspAla 354
Db 1369 CACACCTTCGACGTCAGCGCGCGCACCTTCGTGACCGCCACCTCTACTGGAC----- 1422
QY 355 ProAlaSerThrSerAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAla 374
Db 1423 -----ACGGGCTCGAGCGACATCGACCTCTACTCTACGAC 1458
QY 375 ProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAsp 394
Db 1459 CCCAACGGGAACGAG---GTTGACTACTCTTACACCGGCTACTAC----- 1500
QY 395 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrVal 414
Db 1501 -----GGCTTCGAGAGGTGCGGTACTACACCGACCGCGCGAACCCTGACGCTC 1551
QY 415 GluValGlnAlaTyrAsnValProGlnGlyProGlnAlaPheSerLeuAlaIleValAsn 434
Db 1552 AAGGTGTCGACGTAC-----AAGGCGCGCGCGAAGTACCAGGTGCGACGTCTGACG 1602

RESULT 6
US-09-445-472-11
; Sequence 11, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikuroshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
```

```

; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151989/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-11

Alignment Scores:
Pred. No.: 4,81e-37 Length: 1977
Score: 450.50 Matches: 138
Percent Similarity: 47.05% Conservative: 69
Best Local Similarity: 31.36% Mismatches: 170
Query Match: 20.02% Indels: 63
DB: Gaps: 15

US-09-985-689A-6 (1-434) x US-09-445-472-11 (1-1977)
QY 8 VallysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnValValala 27
Db 433 ATAGGGCGGATACCGTCTCGAAGCTCTCCGCTACGACGAGCGGTGTGTGGTGGC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATCTGTCGATACGGGTATAGACGCGAAC-----CACCCGATCTGAAG 534
QY 48 GlyIleThrAlaIleTyr---AlaLeuGlyArgThrAsnAlaAsnAspProAsn 66
Db 535 GCGAAGGTCATAGGTGTCGACGCGCTCAACGGCAGTTCGACCCCTCAGTACGACG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThrSerAsnLys----- 84
Db 595 GACACGCGGACCCACGTTGCGGGTATCGTTCGCGAACCGCGGTTAACTCCAGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMet-----AspSerAsn 101
Db 655 ATAGCGGTGCGCCCGCGCGAAGCTCTCGCGGTCAAGGTTCTCGGTGCGCGACGTTCC 714
QY 102 GlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSerAla 121
Db 715 GGAAGCGTCTCCACCATCATCGCGGTGTGTGCTGGTCCGACGACGACCAAGTAC 774
QY 122 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThrAsp 141
Db 775 GGGATAAGGGTCATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGACGCGAACCGACTCC 834
QY 142 SerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAlaGly 161
Db 835 CTCAGTCAGCGCGTCAACACCGCTGGGACGCGGTATAGTAGTCTCGGTCCGCGCGC 894
QY 162 AsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 181
Db 895 AACAGCGCGGCGAACAACCTACACCGTCGGCTCACCCTCGCGCGCGCGGACGAGTCA 954
QY 182 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 201
Db 955 GTCGGTGCA-----GTTGACAGCAACGACAAC 981
QY 202 ValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspValMet 221
Db 982 ATCCGACGCTTCTCCAGCAGGAGGACCGACCGCGGACGGAAGGCTCAAGCCGAGTCT 1041
QY 222 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPheTrp 241
Db 1042 GCCCGCGCGGTGACATCATAGCCCGCGCGAGC-----GGAACCAAGTGGGC 1092
QY 242 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 261
```

```

Db 1093 ACCCGATAAGCACTACTACCAAGCGCTCTGGAAACCATGCGCCACCCCGACGTT 1152
Qy 262 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleTysAsnArgGlyIleThrPro--- 280
Db 1153 TCGGCGTGTGGCGCTCATCTCCAG-----GCCACCGGAGTGCGACCCCGGAC 1203
Qy 281 -----LysProSerLeuLeuLysAlaLeuIleAlaGlyAlaThrAspIleGly--- 297
Db 1204 AAGGTGAAGACGCGCTCATCGAGACCGCGCATAGTCGCGCCCAAGGAGATAGCGGAC 1263
Qy 298 LeuGlyTyrProSerGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeu--- 316
Db 1264 ATCGCTAC-----GGTGGGTAGGGTGAACGCTCTACAAGGCCCATCAAG 1308
Qy 317 -----AsnValAlaPheValAsnGlnThrSerSerLeuSerThrAsnGlnLysAlaThr 334
Db 1309 TACGACGACTACGCCAAGCTCACCTTCACCGGCTCCGTCGCGGACAAAGGAGCGCCACC 1368
Qy 335 TyrSerPheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValTyrSerAspAla 354
Db 1369 CACACCTTCAGCTCAGCGCGCCACCTTCGAGCGCGCCACCTCTACTGGGAC----- 1422
Qy 355 ProAlaSerThrSerAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAla 374
Db 1423 -----ACGGGCTCGAGCGACATCGACCTCTACTCTACGAC 1458
Qy 375 ProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAspAsnAsnTyrAsp 394
Db 1459 CCGAACGGGAAGCAG---GTTGACTACTCTACACCGGCTACTAC----- 1500
Qy 395 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrVal 414
Db 1501 -----GGTTCGAGAAGTGGCTACTACAACCGCGACCGCGGAACCTGCGCGTC 1551
Qy 415 GluValGlnAlaTyrAsnValProGlnGlyProGlnAlaPheSerLeuAlaIleValAsn 434
Db 1552 AAGGTCTCAGCTAC-----AAGGCGCGCGGAACCTACCAAGCTCGACGTCGTGACG 1602

```

RESULT 7

```

US-09-445-472-2
; Sequence 2, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Kunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-2

```

```

Alignment Scores:
Pred. No.: 8, 81e-33
Score: 406.50
Percent Similarity: 44.22%
Best Local Similarity: 30.22%
Query Match: 82
DB: 17

```

```

US-09-985-689A-6 (1-434) x US-09-445-472-2 (1-1236)
Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyCl 24
Db 24 GTCTCAGCTCAAGTTTATGCGCACTTACCTTGGAACTTGGATATGATGGTTCTGGAAT 83
Qy 24 nValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 84 CACAATAGCAATTAATGACACTGGAATTGAC-----GCITCTCATCC 125
Qy 44 uAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 126 AGATCTCCCAAGGAAAAGTA-----ATTGGGTGGGTAGATTGTTGTCATGCG 170
Qy 64 -----AspProAsnGlnHisGlyThrHisValAlaGlySerValLeuCl 78
Db 171 TAGAGATTATCCATGATGACCATGACATGGAACATCATGTAGCTTCAATAGCAGCTGG 230
Qy 78 YAsnGlyThrSerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGGAGCAGCAAGTAATGCGCAAGTACAAAGGAATGGCTCCAGGAGCTAAGCTGCGGG 290
Qy 94 eGlnSerValMet-----AspSerAsnGlyGlyLeuGlyGlyLeuProSerAsnValSe 112
Db 291 AATTAGGTTCTAGGTGCGCATGGTTCTGGAACATATCTACTATATTAATTAAGGAGTTGA 350
Qy 112 rThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAl 132
Db 351 GTGGCGGCTTGATACAAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTCTGCTTC 410
Qy 132 aProValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAs 152
Db 411 AAGCCGAGCTCAGATGGTACTGACGCTTAAGTCAGGCTGTTAATGCGAGCTGGGATGC 470
Qy 152 nAspMetAlaValLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAl 172
Db 471 TGGATTAGTTGTTGTTGTTGCTGGAAACAGTGGACCTAACCAAGTATACAACTCGTTC 530
Qy 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh 192
Db 531 TCCAGCAGCTGCAAGCAAAAGTTATTACAGTTGGAGCC----- 567
Qy 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLy 212
Db 568 -----GTTGACAGTATGATGTTATTAACAAGCTTCTCAAGCAGAGGCGCACTGC 617
Qy 212 sAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSe 232
Db 618 AGACGCGAGCTTAAGCCTCAGGTGTTGCTCCAGGAACTGGATAATTGCTGCCAGAGC 677
Qy 232 rSerLeuAlaProAspSerSerPheTyrAlaAsnHisAspSerLysTyrAlaTyrMetG1 252
Db 678 AAGT-----GGAACTAGCATGGGTCAACCAATTAATGACTATTACACAGCAGCTCC 728
Qy 252 yGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPh 272
Db 729 TGGGACATCAATGGCAACTCTCAGCTAGCTGTTGTTGTCAGGAACTGCTTGTCTCAA-- 783
Qy 272 eIleLysAsnArgGlyIleThrPro-----LysProSerLeuLysAlaAlaLe 289
Db 784 ----GCACCGCGAGCTGGAGCTCCAGACAAAGTAAACACAGCCCTCATAGAACTGCTGA 839
Qy 289 uIleAlaGlyAlaThrAspIleGly---LeuGlyTyrProSerGlyAsnGlnGlyTyrpG1 308
Db 840 TATCGTAAAGCCAGATGAATAGCCGATATAGCCTAC-----GGTGCGAG 884
Qy 308 yArgValThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSe 325
Db 885 TAGGTTAATGATACAAAGCTATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 944
Qy 325 rSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAlaGlnSerGlyLysProLe 345
Db 945 ATATGTTGCCACAAAGGCGAGCCAACTCCACGAGTTCGTTATTAGCGGAGCTTCGTTCTG 1004

```

[illegible]

Db 840 TATCGTAAAGCCAGATGAATAGCCGATATAGCCTAC - - - - - GGTGCAGG 884
Qy 308 YargValThrLeuAspLysSerLeu - - - - - AsnValAlaPheValaenGluThrSe 325
Db 885 TAGGTTAATGCATCAAGGCTATAAAGTACGATACGATCAAGCTAGTGTCTCAGTGG 944
Qy 325 rSerLeuSerThrAsnGlnLysAlaThrTy-SerPheThrAlaGlnSerGlyLysProLe 345
Db 945 ATATGTTGCCAACAAAGGAGCCAAACTCACCAGTTGCTTATTAGCGGAGCTTCGTTGCT 1004
Qy 345 uLysLeSerLeuValThrSerAspAlaProAlaSerThrSerAlaSerValThrLeuVa 365
Db 1005 AACTGCCACATTATATCGGCAATGCCAAT - - - - - 1035
Qy 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyValGlyAsnAspPh 385
Db 1036 -AGCGACCTTGATCTTACCTCTACGATCCCAATGGAACACAG--GTTGACTACTCTTA 1091
Qy 385 eThrAlaProTyAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1092 CACCGCCTACTAT - - - - - GGATTGCAAAAGGTTGTTATTA 1127
Qy 405 nAlaProGlnSerGlyThrTyThrValGluValGlnAlaTyAsnValProGlnGlyPr 425
Db 1128 CAAACCACTGATGGAACATGACAAATTAAGTTGTAAGCTACAGC - - - - - GGAAG 1178
Qy 425 oGlnAlaPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTATCAAGTAGATGTGTAAGT 1206

RESULT 9

US-08-894-818B-34
Sequence 34, Application US/088994818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPER-THERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIORITY APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197

TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 1962 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-34
Alignment Scores:
Pred. No.: 1,846-32 Length: 1962
Score: 406.50 Matches: 136
Percent Similarity: 44.22% Conservative: 63
Best Local Similarity: 30.22% Mismatches: 169
Query Match: 18.07% Indels: 82
DB: 3 Gaps: 17
US-09-985-689A-6 (1-434) x US-08-894-818B-34 (1-1962)
Qy 12 ValAlaGlnSerSerTyrgly-LeuTy - - - - - GlyGlnGlyCl 24
Db 420 GTCTGCAGCTCAAGTTATGCGCACTTACGTTTGGAACTTGGGATATGATGGTTCTTGAAT 479
Qy 24 nValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMethHisG1 44
Db 480 CACAATAGGATTAATTGACACTGGAATTGAC - - - - - GCITTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaIleTyAlaLeuGlyArgThrAsnAsnAlaSer-- 63
Db 522 AGATCTCCAAAGGAAAAGTA - - - - - ATTGGTGGGTAGATTTTGTCAATGG 566
Qy 64 - - - - - AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 567 TAGGATTATCCATACGATGACCATCGAATGAACTCATGTAGTTCATAGCAGCTGG 626
Qy 78 YAsnGlyThrSerAsn - - - - - LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGGCAAGTACAAAGGGAATGGCTCCAGGAGCTAAGCTGCGG 686
Qy 94 eGlnSerValMet - - - - - AspSerAsnGlyGlyLeuGlyGlyLeuProSerAsnValSe 112
Db 687 AATTAAGTTCTAGTGCCGATGGTCTTGGAGCATATCTACTATATTAATTAAGGGAGTTGA 746
Qy 112 rThrLeuPheSerGlnAlaTySerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGCGGCTTGATAACAAAGATAAGTACGGAATTAAGGTCATTAATCTTCTCTGTTTC 806
Qy 132 aProValAsnGlyAlaTyThrThrAspSerArgAsnValAspAspTyValArgLysAs 152
Db 807 AAGCCAGAGCTCAGATGCTGCTCAGCGCTTAAGTCAGGCTGTTAATGCGAGCGGATGC 866
Qy 152 nAspMetAlaValLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAl 172
Db 867 TGGATTAGTTGTTGTTGCTTCCGCTGGAAACAGTGGACCTAACCAAGTATACAAATCGTTC 926
Qy 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh 192
Db 927 TCCAGCAGCTGCAAGCAAAAGTTATTACAGTTGAGGCC - - - - - 963
Qy 192 eGlySerTyAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLy 212
Db 964 - - - - - GTTGACAAAGTATGATGTATACAGCTTCTCAAGCAGAGGCGCAACTGC 1013
Qy 212 sAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSe 232
Db 1014 AGCGGAGGCTTAAGCTGAGTTGTTGCTCCAGGAAATCGATTAATGCTGCCAGAGC 1073
Qy 232 rSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyAlaTyMetG1 252
Db 1074 AAGT - - - - - GGAATGACATGGGTCAACCAATTAATGACTATACAGCAGCTCC 1124
Qy 252 yGlyThrSerMetAlaThrProfileValAlaGlyAsnValAlaGlnLeuArgGluHisPh 272

Db 1125 TGGGACATCAATGGCACTCTCACGTAGCTGGTATTGCGAGCCCTCTGCTCCAA----- 1179
Qy 272 eilelyAsnArgGlyIleThrPro-----LysProSerLeuLeuLysAlaAlaLe 289
Db 1180 -----GCACACCGAGCTGACTCCACACAAAGTAAACAGAGCCCTCATAGAACTGCTGA 1235
Qy 289 uileAlaGlyAlaThrAspIleGly---LeuGlyTyrProSerGlyAsnGlnGlyTyrPgl 308
Db 1236 TATCGTAAAGCCAGATGAATAGCCGATATAGCTAC-----GGTGCAGG 1280
Qy 308 YargValThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSe 325
Db 1281 TAGGGTTAATGATACAGCTATAAAGTACAGTAACTATGCAAGCTAGTCTTCACTGG 1340
Qy 325 rSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAlaGlnSerGlyLysProLe 345
Db 1341 ATATGTTGCCAACAAAGGAGCCAAACTCACCAGTTCGTTATTAGCGGAGCTTCGTTGCT 1400
Qy 345 uilylleserLeuValTyrSerAspAlaProAlaSerThrSerAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATTACTGGGACCAATGCCAAT----- 1431
Qy 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPh 385
Db 1432 -ACGACCTTGACTTTACCTCTACGATCCCAATGGAACACAG---GTTGACTACTCTTA 1487
Qy 385 eThrAlaProTyrAspAsnAsnTyrPaspGlyArgAsnValGluAsnValPheIleAs 405
Db 1488 CACCGCCTACTAT-----GGATTGCAAAAGGTTGGTTATTA 1523
Qy 405 nAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsnValProGlnGlyPr 425
Db 1524 CAACCCCACTGATGGACATGGACAATTAAGGTTGTAAGCTACAGC-----GGAAG 1574
Qy 425 oGlnAlaPheSerLeuAlaIleValAsn 434
Db 1575 TGCAAACTATCAAGTAGATGGTAAAGT 1602

RESULT 10

US-09-445-472-15
; Sequence 15 Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-15

Alignment Scores:

Pred. No.:	1,84e-32	Length:	1962
Score:	406.50	Matches:	136
Percent Similarity:	44.22%	Conservative:	63
Best Local Similarity:	30.22%	Mismatches:	169
Query Match:	18.07%	Indels:	82
DB:	4	Gaps:	17

US-09-985-689A-6 (1-434) x US-09-445-472-15 (1-1962)
Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyGl 24
Db 420 GTCTGCAGCTCAAGTTATGCAACTTACGTTGGAACTTGGGATATGATGTTCTCGAAT 479
Qy 24 nValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGl 44
Db 480 CACAATAGGAATTAATGACACACTGGAATTGAC-----GCTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAAGGAAAAGTA-----ATTGGTGGGTAGATTTTGTCAATGG 566
Qy 64 -----AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db 567 TAGGAGTTATCCATPACCATGACATGACATGGAATCATGTAGCTTCAATAGCAGCTGG 626
Qy 78 YAsnGlyThrSerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGCAAGTACAGGAATGGCTCCAGGAGCTAAGCTGGCGG 686
Qy 94 eGlnSerValMet-----AspSerAsnGlyGlyLeuGlyGlyLeuProSerAsnValSe 112
Db 687 AATTAAAGTTTCTAGTCCCGATGGTTCTGGAAGCATATCTACTATAATTAAGGAGTTGA 746
Qy 112 rThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAl 132
Db 747 GTGGCGCGTTCATACAAAGATAGTACGGAATTAAGTTCATTAATCTTCTCTGTTTC 806
Qy 132 aProValAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyrValargLysAs 152
Db 807 AAGCAGAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTTAATGACGCGTGGGATGC 866
Qy 152 nAspMetAlaValLeuPheAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAl 172
Db 867 TGGATTAGTTGTTGTTGCTGCGCTGGAACAGTGGACCTTACAGAGTATACATCGTTTC 926
Qy 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh 192
Db 927 TCCAGCAGCTGCAAGCAAGTTATTACAGTTGGAGCC----- 963
Qy 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLy 212
Db 964 -----GTTGACAGTATGATGATTATACAGCTTCTCAGCAGAGGGCCACTGC 1013
Qy 212 sAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSe 232
Db 1014 AGACGGCAGGCTTAAGCCTGAGTTGTGCTCCAGGAAACTGGATAATGTCGCCAGC 1073
Qy 232 rSerLeuAlaProAspSerSerPheTyrAlaAsnHisAspSerLysTyrAlaTyrMetGl 252
Db 1074 AAGT-----GGAAGTACATGGGTCAACCAANTTAATGACTATTACAGCAGCTCC 1124
Qy 252 yGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPh 272
Db 1125 TGGGACATCAATGGCAACTCTCTCAGTAGTGGTATTGTCAGCCCTCTTGCTCCAA----- 1179
Qy 272 eilelyAsnArgGlyIleThrPro-----LysProSerLeuLysAlaAlaLe 289
Db 1180 -----GCACACCGAGCTGAGCTCCAGACAAAGTAAACACAGCCCTCATAGAACTGCTGA 1235
Qy 289 uileAlaGlyAlaThrAspIleGly---LeuGlyTyrProSerGlyAsnGlnGlyTyrPgl 308
Db 1236 TATCGTAAAGCCAGATGAATAGCCGATATAGCTAC-----GGTGCAGG 1280
Qy 308 YargValThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSe 325
Db 1281 TAGGGTTAATGATACAGGCTATAAAGTACAGTAACTATGCAAGCTAGTCTTCACTGG 1340
Qy 325 rSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAlaGlnSerGlyLysProLe 345
Db 1341 ATATGTTGCCAACAAAGGAGCCAAACTCACCAGTTCGTTATTAGCGGAGCTTCGTTGCT 1400

QY 345 ulysileSerLeuValTrpSerAspAlaProAlaSerThrSerAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATTATCTGGACAATGCCAAT----- 1431
QY 365 lAsnAspLeuValIleThrAlaProAsnGlyThrIysTyrValGlyAsnAspPh 385
Db 1432 -AGCGACCTTGATCTTACCTCTACGATCCCAATGGAACACAG--GTTGACTACTCTTA 1487
QY 385 eThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1488 CACCGCCCTACTAT-----GGAATCGAAAGGTTGGTTATTA 1523
QY 405 nAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsnValProGlnGlyPr 425
Db 1524 CAACCCAACTGATGAACATGACAAATTAAGTTGTAAGCTACAGC-----GGAAG 1574
QY 425 oglnAlaPheSerLeuAlaIleValAsn 434
Db 1575 TCGAACTATCAAGTAGATGTTGTAAGT 1602

RESULT 11

US-08-894-818B-6
Sequence 6, Application US/08894818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1977 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-894-818B-6

Alignment Scores: 2.37e-32 Length: 1977
Pred. No.: 405.50 Matches: 134
Score: 45.11% Conservative: 69
Percent Similarity: 29.78% Mismatches: 165
Best Local Similarity: 18.02% Indels: 82
Query Match: 3 Gaps: 17
DB: 17
US-09-985-689A-6 (1-434) x US-08-894-818B-6 (1-1977)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 420 GTCTGCAGCTCAAGTTATGCAACTTACGTTTGGAACTTGGGATATGATGTTCTGGAAT 479
QY 24 nValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 480 CACAATAGGAATAATTGACACTGGAATTCAC-----GCTTCTCATCC 521
QY 44 uAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAAAGGAAGAAGTA-----ATTGGTGGGTAGATTTGTCAATGG 566
QY 64 -----AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 567 TAGGAGTTATCCATACGATGACCATGACATGGAAGTCACTAGCTTCAATAGCAGCTGG 626
QY 78 YAsnGlyThrSerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGGCAAGTACAAAGGAATGGCTCCAGGAGCTAAGCTGGCGGG 686
QY 94 eGlnSerValMet-----AspSerAsnGlyLeuGlyGlyLeuProSerAsnValse 112
Db 687 AATTAAGGTTCTAGTGCCGATGTTCTGGAAGCATATCTACTATAATTAAGGAGCTGA 746
QY 112 rThrLeuPheSerGlnAlaTyrSerAlaGlyAlaAaGlieHisThrAsnSerTrpGlyAl 132
Db 747 GTGGGCGCTTGATACAAAGATAAGTACGGAATAAGGTCATTAATCTTCTTCTGCTTC 806
QY 132 aProValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAs 152
Db 807 AAGCCAGAGCTCCGACGGAACCGACTCCCTCAGTCAGGCGGTCAACAACGCTGGGACGC 866
QY 152 nAspMetAlaValLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAl 172
Db 867 CGGTATAGTAGTCTGCGTCGCGCGCAACACGCGGCGCAACACCTACACCGCTCGGCTC 926
QY 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh 192
Db 927 ACCGCGCGCGGACGACGAGTCTATACCGTCCGTGCA----- 963
QY 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLy 212
Db 964 -----GTTGACAGCAACACGACACATCGGCAGCTTCTCCAGCAGGCGGACCGCCG 1013
QY 212 sAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgse 232
Db 1014 GGACGGAAGGCTCAAGCGGAAGTGTCTGCGCCCGCGGCTTGACATCATAGCCCCGGCGGC 1073
QY 232 rSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetG1 252
Db 1074 CAGC-----GGAACCAAGCATGGGCACCCCGATAAAGCACTACTACACCAAGGCGCTC 1124
QY 252 yGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPh 272
Db 1125 TGGAAACCAAGCATGGCCACCCCGACGCTTTCGGGCGTTGGCGGCTCATCTCTCCAG----- 1179
QY 272 eIleLysAsnArgGlyIleThrPro-----LysProSerLeuLysAlaAla 289
Db 1180 -----GCCCAACCGAGCTGGACCCCGGACAAGGTGAAGACCGCTCATCGAGACCGCGCA 1235
QY 289 uIleAlaGlyAlaThrAspIleGly---LeuGlyTyrProSerGlyAsnGlnGlyTrpG1 308
Db 1236 CATAGTCGCCCCCAAGGAGATAGCGGACATCGGCTAC-----GGTGGCGGG 1280

```
QY 308 YATGValThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSe 325
Db 1281 TAGGTTGAAGCTTACAGGCCATCAAGTACGACGACTACGCGACCTCACCTTACCGG 1340
QY 325 rSerLeuSerThrAsnGlnLysAlaThrTySerPheThrAlaGlnSerGlyLysProLe 345
Db 1341 CTCGTCGCGCACAAGGAGCGCCACCCACACCTTCGACGTCGAGCGGCCACCTTCTGT 1400
QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrSerAlaSerValThrLeuVa 365
Db 1401 GACCGCCACCTTCTACTGGAC-----ACGGGCTC 1430
QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyTyValGlyAsnAspPh 385
Db 1431 GAGCGACATCGACTCTACCTCTACGACCCCAACGGAACGAG---GTTGACTACTCTCTA 1487
QY 385 eThrAlaProTyAspAsnAsnTrpAspGlyAAsnValGluAsnValPheIleAs 405
Db 1488 CACCGCTTACTAC-----GGCTTCGAGAAGTCGGCTACTA 1523
QY 405 nAlaProGlnSerGlyThrTyThrValGluValGlnAlaTyAsnValProGlnGlyPr 425
Db 1524 CAACCGACCGCGGACCTGAGCGGTCAAGTCTGTCAGCTAC-----AAGGGCGC 1574
QY 425 cGlnAlaPheSerLeuAlaIleValAsn 434
Db 1575 GCGGAACCTACGAGTCGACGTCGTCAGC 1602

RESULT 12
US-09-000-016-3
; Sequence 3, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2539 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
```

```
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces viridosporus
; STRAIN: A-914
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
US-09-000-016-3
```

```
Alignment Scores:
Pred. No.: 1.74e-27 Length: 2539
Score: 360.50 Matches: 146
Percent Similarity: 43.38% Conservative: 57
Best Local Similarity: 31.20% Mismatches: 175
Query Match: 16.02% Indels: 90
DB: 3 Gaps: 22
```

US-09-985-689A-6 (1-434) x US-09-000-016-3 (1-2539)

```
QY 2 AspValAlaAaGGGlyIleValLysAlaAspValAlaGlnSerSerTyGlyLeuTyGly 21
Db 953 GACACGTCCTCGCGGACATCGGCGCCCAAGCGTGTCTCCGCC---GGCTACGACGGC 1009
QY 22 GlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
Db 1010 AAGGCGGTGAAGATCGCGCTCTCGACACCGGTCTCGACACGAGC----- 1054
QY 42 MethHisGluAlaPheArgGlyIleThrAlaIleTyAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGGACCTGAAGGGCGGGTGACCGGCTCAAGAACTTCACCGCGCCCGCC 1111
QY 62 AlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThr 81
Db 1112 GCGCGGACAAAGTGGGCGCACCGCACCGTCGCTCGATCGCGGGGCGACGGCGCC 1171
QY 82 SerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerVal 97
Db 1172 CAGTCCAAAGGCAAGTACAAAGGGGTGCGACCCCGCGCGCGATCTCTCAACGGCAAGTC 1231
QY 98 MetAspSerAsnGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGln 117
Db 1232 CTCGACGACTCGGT-----TTGGGCGACACTCCGCGCATCTCCGCGCATGAGTGG 1285
QY 118 AlaTySerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAla 137
Db 1286 GCGGCGCGGCGAGCGCGAGCTGTCACCATGAGCCTCGGC-----GGCATG 1333
QY 138 TyrThrThrAspSerArgAsnValAspTyTyValArgLys-----AsnAspMetAla 155
Db 1334 GACACACCGGAGACCGCCGCTGAGCGCGGTGCAACAGCTGTCGCGCGAGAGGGC 1393
QY 156 ValLeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaPro 173
Db 1394 GTCTTGTTCGCTATCGCGCGCGCAACGAGGGCGCGAG-----TCGATCGTTCGGCC 1447
QY 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GGCAGCGCGGACCGCGCTCTACCGTCGGCGCC----- 1480
QY 194 SerTyAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLys 212
Db 1481 -----GTCGACGACAAAGCAAGACTCCCGACTTCTCTCCACCGCGCCCGCTCGGC 1534
QY 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
Db 1535 GACGCGCCATCAAGCGGACGTCACCGCTCCCGCGTGGACATCACCGCGCTCGCGG 1594
QY 233 SerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyAlaTyMetGly 252
Db 1595 GAGGGCAACGACATCGCGCAGGAGTGGTGAGGGACCGCGCTACATGACCATCTCC 1654
QY 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgLuhisPhe 272
```


1655 GGCACGTCGATGGCGACCCCGCAGTCGCGGGCGCGGCGCCCTCTCTGAGCAGCAG--- 1711
Qy 273 IleIysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
Db 1712 -----CACCCCGACTGGACCTCCGCCCAACTGAAGGCGCG 1747
Qy 290 IleAlaGlyAlaThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTrpGlyArg 309
Db 1748 CTCACCGCTCCACCAAG---GGCGGAAGTACACCCGTTTCAGCAGGTTTCGGGCCGG 1804
Qy 310 ValThrLeuAspLysSerLeuAsnValAlaPheValAsnGluThrSerLeuSer--- 328
Db 1805 ATCCAGGCCGACAAAGGCTCCAGCAGACCGTGTATCGCCAGCCCGTCTCGGTGAGCTTC 1864
Qy 329 -----ThrAsnGlnLysAlaThrTyrSer 336
Db 1865 GGGTCCAGCAGTGGCGGCACACGACGACGACCGGTACCAAGCAGCTGACCTACCGC 1924
Qy 337 PheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1925 AACCTGGCACCAGGACGTCACGCTGAAGCTCAGCTGACCGCCACCGCCCAAGGCG 1984
Qy 354 -----AlaProLaser-----ThrSerAlaSerValThrLeuValAsnAspLeuAsp 369
Db 1985 AAGCGCGCCCGCGCGCTTCTTACGCTGGCGGCCACACG----- 2026
Qy 370 LeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyr 389
Db 2027 -----GTGACCGTCCCGCGGCGGCGAGC---GCCTCCGTCCACATGACCGCC----- 2071
Qy 390 AspAsnAsnTrpAspGlyArg-----AsnAsnValGluAsnValPheIleAsnAlaProGln 408
Db 2072 GACACCGCGTCCGCGCACGGTGGACGGCGCTACTCGCGTACCTGGTTCGCG-----ACG 2128
Qy 409 SerGlyThrTyrThrVal-----GluValGlnAlaTyrAsn 420
Db 2129 GCGCGCGGCGAGCGTCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2188
Qy 421 Val-ProGlnGlyProGlnAla 427
Db 2189 GTGACCGTCCGCGACATCGGCC 2210

RESULT 13
US-09-514-340-3
; Sequence 3, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: Akita ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,340
; FILING DATE: 28-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/000,016
; FILING DATE: January 30, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng

REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-514-340-3

Alignment Scores:
Pred. No.: 1.74e-27 Length: 2539
Score: 360.50 Matches: 146
Percent Similarity: 43.38% Conservative: 57
Best Local Similarity: 31.20% Mismatches: 175
Query Match: 16.02% Indels: 90
DB: 4 Gaps: 22

US-09-985-689A-6 (1-434) x US-09-514-340-3 (1-2539)
Qy 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnInserserTyrGlyLeuTyrGly 21
Db 953 GACACGTCCTCGCGGACATCGCGCCCGCCCAAGCGGTGTCGCGC---GGCTACGACGC 1009
Qy 22 GlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
Db 1010 AAGGCGGTGAAGATCGCGCTCTCGACACCGGTGTGCACACGAGC----- 1054
Qy 42 MetHisGluAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGACCTGAAGGCGCGGTGACCGCTCCAGACCTCACCAGCGCGCGCGCG 1111
Qy 62 AlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThr 81
Db 1112 GCGCGCGCAAGGTGGCGCACCGCACCGCCTCGATCGCGCGCGCGCGCGCGCGCG 1171
Qy 82 SerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerVal 97
Db 1172 CAGTCCAAAGGCAAGTACAGGGCGTGCACCCCGCGCGCGCTCCCAACGCGCAAGGTC 1231
Qy 98 MetAspSerAsnGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGln 117
Db 1232 CTCGACGACTCCGCT-----TTCGGCGACGACTCCCGCATCTCCCGCGCATGAGTGG 1285
Qy 118 AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAla 137
Db 1286 GCGCGCGCGCGCGCGCGCGCGCTCGTACCATGAGCTGGGC-----GGCATG 1333
Qy 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgLys-----AsnAspMetAla 155
Db 1334 GACACCGGAGACCGCCCGCTGGAGGCGCGCTGCACAGCTGTCGCGCGAGAGGCGC 1393
Qy 156 ValLeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaPro 173
Db 1394 GTCTGTGTCG 1447
Qy 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GCGAGCGCGCGCGCGCGCGCGCTCAGCTCGCGCGC----- 1480

```
Qy 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArglyPro----ThrLys 212
Db 1481 -----GTCGACGACAGGACAACTCGCGGACTTCTCTCCACCGGCCCCCGCTCGGC 1534
Qy 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaAgsr 232
Db 1535 GACGGCGCCATCAAGCGGAGCTCACCGCTCCGCGCTGGACATCAGCGCGCTCGGC 1594
Qy 233 SerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGly 252
Db 1595 GAGGGCAACGACATCGCGCAGGAGTGGTGAGGAGCGCGCGCTACATGACCATCTCC 1654
Qy 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
Db 1655 GGCAGCTCATGGGACCCCGACAGTGGGGCGCGCGCGCTCTCTGAAGCAGCAG--- 1711
Qy 273 IleLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
Db 1712 -----CACCCCGACTGGACCTCGCGCGAAGTGAAGGGCGCG 1747
Qy 290 IleAlaGlyAlaThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTyrGlyArg 309
Db 1748 CTCACCGGCTCCACCAAG---GGGGCAAGTACACCCCGTTCGAGCAGGTTGGGGCGG 1804
Qy 310 ValThrLeuAspLysSerLeuAsnValAlaPheValAsnGlnThrSerSerLeuSer--- 328
Db 1805 ATCCAGGCGGACAGGCGCTCCAGCAGACCGTGTATCGCGACCGCGTCTCGTGAGCTTC 1864
Qy 329 -----ThrAsnGlnLysAlaThrTyrSer 336
Db 1865 GCGGTCCAGAGTGGCGGCACACGACGACGCGGTTCACACAGCAGCTGACCTACCGC 1924
Qy 337 PheThrAlaGlnSerGlyLysPheProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1925 AACCTCGGACCCAGGACGTCAGCTGAGTGCAGTGCAGCGCGCACCGACCCCAAGGGC 1984
Qy 354 -----AlaProAlaSer-----ThrSerAlaSerValThrLeuValAsnAspLeuAsp 369
Db 1985 AAGGCGCGCCCGCGCGGCTTCTTCACGTGGCGGCACACG----- 2026
Qy 370 LeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyr 389
Db 2027 -----GTACCGTCCCGCGCGCGCGCAGC---GCCTCCGTGCATCAGCGCC----- 2071
Qy 390 AspAsnAsnTrpAspGlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGln 408
Db 2072 GACACCGCGCTCGCGCGCACGCGTGGACGCGCGGTACTCGCGGTACGTGTCGCC---ACG 2128
Qy 409 SerGlyThrTyrThrVal-----GluValGlnAlaTyrAsn 420
Db 2129 GCGCGCGGCGACAGCGTCCGACGCGCGCGCGGTGTCAGCGCGAGTGCAGTACGAC 2188
Qy 421 Val-ProGlnGlyProGlnAla 427
Db 2189 GTGACCGTCCGGCACATCGGCC 2210
RESULT 14
US-09-000-016-1
; Sequence 1, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces viridosporus
; STRAIN: A-914
; ORGANISM: Streptomyces antibioticus
; STRAIN:
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2540...2809
; IDENTIFICATION METHOD: P
;
; US-09-000-016-1
;
; Alignment Scores:
; Pred. No.: 6,77e-27 Length: 2809
; Score: 355.50 Matches: 145
; Percent Similarity: 43.16% Conservative: 57
; Best Local Similarity: 30.98% Mismatches: 176
; Query Match: 15.80% Indels: 90
; DB: Gaps: 22
;
; US-09-985-689A-6 (1-434) x US-09-000-016-1 (1-2809)
;
; Qy 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
; Db 953 GACACGTCCGTGGGAGATCGCGCTCTCTGGACACCGGTGTGCACACGAGC---GGCTACGACGCGC 1009
;
; Qy 22 GlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
; Db 1010 AAGGGCGTGAAGATCGCGTCTCTGGACACCGGTGTGCACACGAGC----- 1054
;
; Qy 42 MethHisGluAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsn 61
; Db 1055 ---CATCCGACCTGAAGGCGCGGGTACCGCTCCAGACTTACCCGCGCGCGCGC 1111
;
; Qy 62 AlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThr 81
; Db 1112 GCCGGCGACAGGTGGCGCCACGCGCACCGACCGCTCGCTCGATCGCGCGCGCGCGCGC 1171
;
; Qy 82 SerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerVal 97
; Db 1172 CAGTCCAGGCGCAAGTAGTAAGGGCGTCTGACCCCGCGCGCGCGCTCTCAACAGCGAAGGTC 1231
```

QY 98 MetAspSerAsnGlyGlyLeuProSerAsnValSerThrLeuPheSerGln 117
Db 1232 CTCGACGACTCCGGT-----TTCGGCGACGACTCCGGCATCTCCGGCGCATGAGTGG 1285
QY 118 AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAla 137
Db 1286 GCGCGCGCGAGGCGCGGCGTCAACATGAGCTGGGC-----GGCATG 1333
QY 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgIle-----AsnAspMetAla 155
Db 1334 GACACCGGAGACCGACCGCGTGGAGGGCGGCTGCAACAGTGTCCCGCCGAGAGGCG 1393
QY 156 ValLeuPhe-----AlaAlaGlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaPro 173
Db 1394 GTCCTGTTCGCTATCGCGCGCGCGCAACAGAGGCGCGAG-----TCGATCGTTCGCC 1447
QY 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GGCAGCGCGGACCGCGCTCCACCGTCCGCGCGC----- 1480
QY 194 SerTyrAlaAspAsnIleAsnHisAlaGlnPheSerSerArgGlyPro---ThrLys 212
Db 1481 -----GTCGACGACAGGACAGCTCGCGACTTCTCTCCACCGCGCGCGCTCCGCG 1534
QY 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
Db 1535 GACGGCGCCATCAAGCGGACGTCACCGTCCCGGCGGTGACATCACGCGCGCTCGCG 1594
QY 233 SerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGly 252
Db 1595 GAGGGCAACGACATCGCGCGAGGAGTGGTGGAGGCGCGCGCTACATGACCATCTCC 1654
QY 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
Db 1655 GGCACGTCGATGGGACCGCGCGTCCGCGCGCGCGCTCTCTGAGCAGCAGCAG--- 1711
QY 273 IleLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
Db 1712 -----CACCCGACCTGGACCTCCCGCGCGCGCTGAGGCGCG 1747
QY 290 IleAlaGlyAlaThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTrpGlyArg 309
Db 1748 CTCACCGCTCCACCAAG---GGCGGCAAGTACACCCCGTTCGAGCAGGCTTCGGGCGG 1804
QY 310 ValThrLeuAspLysSerLeuAsnValAlaPheValAsnGlnThrSerSerLeuSer--- 328
Db 1805 ATCCAGCGCGAAGGCGCTCCAGCAGCGGTGATCCCGCGCGCTCTCGTGAGCTTC 1864
QY 329 -----ThrAsnGlnLysAlaThrTyrSer 336
Db 1865 GCGCTCCAGCAGTGGCGGCACCGACGACGCGCGTCCACCAAGCAGCTGACCTACCG 1924
QY 337 PheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1925 AACCTCGGACCCAGGAGCTCACGCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1984
QY 354 -----AlaProAlaSer-----ThrSerAlaSerValThrLeuValAsnAspLeuAsp 369
Db 1985 AAGCGCGCGCGCGGCTTCTTACGCTGGCGCGCGCGCG----- 2026
QY 370 LeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyr 389
Db 2027 -----GTGACCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2071
QY 390 AsnAsnAsnTrpAspGlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGln 408
Db 2072 GACACCG 2128
QY 409 SerGlyThrTyrThrVal-----GluValGlnAlaTyrAsn 420
Db 2129 GCG 2188
QY 421 Val-ProGlnGlyProGlnAla 427

Db 2189 GTGACCGTCCGGCAGCATCGGCC 2210
RESULT 15
US-09-514-340-1
; Sequence 1, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: Akita ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE 1
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,340
; FILING DATE: 28-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/000,016
; FILING DATE: January 30, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces antibioticus
; STRAIN: <Unknown>
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338..2539
; IDENTIFICATION METHOD: E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2540...2809
; IDENTIFICATION METHOD: P
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-514-340-1
Alignment Scores: 6.77e-27 Length: 2809
Pred. No.: 355.50 Matches: 145
Score: 43.16% Conservative: 57
Percent Similarity: 30.98% Mismatches: 176
Best Local Similarity: 15.80% Indels: 90
Query Match: 4 Gaps: 22
DB:
US-09-985-689A-6 (1-434) x US-09-514-340-1 (1-2809)
Qy 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21

Db 953 GACACGTCCTCGGCGCAGATCGGCGCCCCCAAGCGGTGTGTCGCC---GGCTACGACGGC 1009
Qy 22 GlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
Db 1010 AAGGGCGTGAAGATCGCGCTCTCGACACCGGTGTGACACGAGC----- 1054
Qy 42 MetHisGluAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGACCTTGAAGGGCGGGTACCGCTCCAGAACTTCCACCGCGCGCCCGC 1111
Qy 62 AlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThr 81
Db 1112 GCCCGGACAAAGTGGCGCACCGGACCCACCGTCCGCTCGATCGCGCGGGCACGGGGCC 1171
Qy 82 SerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerVal 97
Db 1172 CAGTCCCAAGGCAAGTACAGGGGTGCGACCGCGCGCGCTCTCAACAGCGCAAGGTC 1231
Qy 98 MetAspSerAsnGlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGln 117
Db 1232 CTCGACGACTCCGGT-----TTCGGCAGCAGCTCCGGCATCTCCGCGCGCATGGAGTGG 1285
Qy 118 AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAla 137
Db 1286 CGGCGCGGCGAGGGCGCGAGCTCGTCAACATGAGCTGGGC-----GGCATG 1333
Qy 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgLys-----AsnAspMetAla 155
Db 1334 GACACCGGAGACCGACCCCGCTGGAGGGCGGGTTCGACAGCTGTCCGCGCGAGAGGGC 1393
Qy 156 ValLeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaPro 173
Db 1394 GTCCGTGTTCCGATCGCGCGCGGCAACGAGGGCGCGAG-----TCGATCGGTTCGCC 1447
Qy 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GCGACGGCGGACGCGCGCTTACCGCTCGCGCC----- 1480
Qy 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLys 212
Db 1481 -----GTCGACGACAGGACAAGCTCGCGGACTTCTCTCCACCGCGCGCGCTCGGC 1534
Qy 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
Db 1535 GACGGCGCATCAAGCGGACGTCCCGCTCCGCGGTGGACATCAGCGCGCGCTCGGCG 1594
Qy 233 SerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGly 252
Db 1595 GAGGGCAACGACATCGGCCAGAGGTGCGTGAGGGACCGCGCGGTACATGACCATCTCC 1654
Qy 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
Db 1655 GGCACGTCGATGGCAGCCCGCACGTGCGGGCGCGCGCCCTCTGAGCAGCAG--- 1711
Qy 273 IleLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
Db 1712 -----CACCCGACTGACCTCCGCGCACTGAAGGGCGG 1747
Qy 290 IleAlaGlyAlaThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTrpGlyArg 309
Db 1748 CTCACCGGTCCACCAAG---GGCGGCAAGTACACCCGCTTCGAGCAGGGTTCGGGCGCG 1804
Qy 310 ValThrLeuAspLysSerLeuAsnValAlaPheValAsnGluThrSerSerLeuSer--- 328
Db 1805 ATCCAGGCCGACAAAGGCGCTCCAGCAGACCGTGTATCGCGACCCCGGTCTCGGTGAGCTTC 1864
Qy 329 -----ThrAsnGlnLysAlaThrTyrSer 336
Db 1865 GCGGTCCAGCAGTGGCGGCACACCGACGAGCGCGTCAACAAAGCAGCTGACCTACCGC 1924
Qy 337 PheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1925 AACCTCGGCACCCAGGACGTACGCTGAGCTGACCGCGCGCGCCCGCCCAAGGGC 1984

Qy 354 -----AlaProAlaSer-----ThrSerAlaSerValThrLeuValAsnAspLeuAsp 369
Db 1985 AAGCGGCGCGCGGGGCTTCTTCCAGCTGGGCGCCACACG----- 2026
Qy 370 LeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyr 389
Db 2027 -----GTGACCGTCCCGGGCGGCGCAGC---GCCTCCGTGACATGACCGCC----- 2071
Qy 390 AspAsnAsnTrpAspGlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGln 408
Db 2072 GACACCCCGCTCGGCGGCACCGTGGACGCGCGGTACTCGGGGTACGTGGTCTGCC---ACG 2128
Qy 409 SerGlyThrTyrThrVal-----GluValGlnAlaTyrAsn 420
Db 2129 GCGCGGCGGACAGCGTCCGACGCGCGCGCGGTGACGCGGAGGTGCGTCTGACGAC 2188
Qy 421 Val-ProGlnGlyProGlnAla 427
Db 2189 GTGACCGTCCGCGCACATCGGCC 2210

Search completed: April 4, 2004, 12:03:34
Job time : 98.5977 secs.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2004, 08:10:21 ; Search time 328.004 Seconds
(without alignments)

4948.852 Million cell updates/sec

Title: US-09-985-689A-6

Perfect score: 2250

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAYNVPGQPAFSLAIWN 434

Scoring table:

BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 2466186 seqs, 1870095128 residues

Total number of hits satisfying chosen parameters: 4932372

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh

-Q=/cg2_1/USPRO.spool/US09985689/runat_31032004_161809_4271/app_query.fasta_1.3498

-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1

-LOOPCL=0 -LOOPEXT=0 -UNITS=bases -START=1 -END=1 -MATRIX=biosum62

-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100

-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0

-MAXLEN=200000000 -USER=US09985689 -CGN 1 1 601 -runat_31032004_161809_4271

-NCFU=6 -ICPU=3 -NO_MMAP -LARGESUBSTR -NEG_SCORES=0 -WAIT -DSFBLCK=100

-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cg2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:

2: /cg2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:

3: /cg2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:

4: /cg2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:

5: /cg2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:

6: /cg2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:

7: /cg2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:

8: /cg2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:

9: /cg2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:

10: /cg2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:

11: /cg2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:

12: /cg2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:

13: /cg2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:

14: /cg2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:

15: /cg2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:

16: /cg2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:

17: /cg2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:

18: /cg2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
------------	-------	-------	--------	----	-------------

1	2143	95.2	1305	15	US-10-385-662-1	Sequence 1, Appli
2	450.5	20.0	1977	13	US-10-090-624-11	Sequence 11, Appli
3	406.5	18.1	1236	13	US-10-090-624-2	Sequence 2, Appli
4	406.5	18.1	1962	13	US-10-090-624-15	Sequence 15, Appli
5	362	16.1	3624	14	US-10-156-761-5701	Sequence 5701, Ap
6	362	16.1	9025608	14	US-10-156-761-1	Sequence 1, Appli
7	355.5	15.8	3417	14	US-10-156-761-3306	Sequence 3306, Ap
8	355.5	15.8	9025608	14	US-10-156-761-1	Sequence 1, Appli
9	339	15.1	135638	14	US-10-314-657-1	Sequence 1334, Ap
10	337.5	15.0	1329	9	US-09-974-300-1934	Sequence 5, Appli
11	317	14.1	4765	13	US-10-090-624-5	Sequence 5, Appli
12	297.5	13.2	1560	15	US-10-084-846A-113	Sequence 113, App
13	297.5	13.2	59816	15	US-10-084-846A-1	Sequence 1, Appli
14	297.5	13.2	59816	15	US-10-084-846A-2	Sequence 2, Appli
15	284.5	12.6	3788	10	US-09-927-827-33	Sequence 33, Appli
16	274	12.2	3743	10	US-09-927-827-29	Sequence 29, Appli
17	272	12.1	1306	9	US-09-966-921A-1	Sequence 1, Appli
18	272	12.1	1330	9	US-09-966-921A-5	Sequence 5, Appli
19	266	11.8	3303	14	US-10-156-761-5384	Sequence 5384, Ap
20	264	11.7	840	14	US-10-209-812-1	Sequence 1, Appli
21	263	11.6	2192	12	US-10-424-599-112429	Sequence 112429, Ap
22	261	11.6	1971	9	US-09-974-300-1935	Sequence 1935, Ap
23	260.5	11.6	2166	12	US-10-344-231-17	Sequence 17, Appli
24	260.5	11.6	2166	12	US-10-363-332A-17	Sequence 17, Appli
25	258	11.5	1485	9	US-09-974-300-1938	Sequence 1938, Ap
26	257	11.4	1140	8	US-08-322-678-11	Sequence 11, Appli
27	257	11.4	1140	16	US-10-323-324-11	Sequence 11, Appli
28	257	11.4	1143	14	US-10-313-853-6	Sequence 6, Appli
29	257	11.4	2588	12	US-10-344-231-20	Sequence 20, Appli
30	257	11.4	2588	12	US-10-363-332A-20	Sequence 20, Appli
31	256	11.4	1140	8	US-08-322-678-12	Sequence 12, Appli
32	256	11.4	1140	16	US-10-323-324-12	Sequence 12, Appli
33	250.5	11.1	1140	15	US-10-146-905A-9	Sequence 9, Appli
34	246.5	11.0	1140	9	US-09-920-118-13	Sequence 13, Appli
35	244.5	10.9	1497	8	US-08-322-678-6	Sequence 6, Appli
36	244.5	10.9	1497	9	US-09-060-854B-1	Sequence 1, Appli
37	244.5	10.9	1497	14	US-10-033-325-1	Sequence 1, Appli
38	244.5	10.9	1497	14	US-10-228-572-1	Sequence 1, Appli
39	244.5	10.9	1497	15	US-10-423-649-1	Sequence 6, Appli
40	244.5	10.9	1497	16	US-10-323-324-6	Sequence 18, Appli
41	244	10.8	2267	12	US-10-344-231-18	Sequence 18, Appli
42	244	10.8	2267	12	US-10-363-332A-18	Sequence 18, Appli
43	242.5	10.8	1149	15	US-10-146-905A-7	Sequence 7, Appli
44	242.5	10.8	1494	14	US-10-104-693-1	Sequence 1, Appli
45	240.5	10.7	1074	9	US-09-837-235-14	Sequence 14, Appli

ALIGNMENTS

RESULT 1
US-10-385-662-1
; Sequence 1, Application US/10385662
; Publication No. US20040002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAITO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAEKI, KATSUHISA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938USO
; CURRENT APPLICATION NUMBER: US/10/385,662
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231

;; PRIOR FILING DATE: 2002-10-18
;; NUMBER OF SEQ ID NOS: 20
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 1
;; LENGTH: 1305
;; TYPE: DNA
;; ORGANISM: Bacillus sp. KSM-KP43
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(1305)
;; OTHER INFORMATION:
US-10-385-662-1

Alignment Scores:
Pred. No.: 7,45e-226 1305
Score: 2143.00
Percent Similarity: 97.93%
Best Local Similarity: 93.55%
Query Match: 95.24%
DB: 15 0

US-09-985-689a-6 (1-434) x US-10-385-662-1 (1-1305)

```
Qy 1 AsnAspValAlaAaGgGlyLeValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1 AATGATGTTGGCGGTGGAAATTTGTCAAAGCGGATGTGGCTCAGACGAGCTACGGGTTGTAT 60
Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGACAAAGACAGATCGTAGCGGTGGCGATACAGGGCTTGATACAGGTGCGCAATGACAGT 120
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsn 60
Db 121 TCGATGCGATGAAGCTTCGCGGGGAAAAATTACTGCATTATATGCAATTTGGGACGGACGAAT 180
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 181 AATGCCAATGATACGAATGCTCATGTACGATGCGATGCGCTCGCTATTAGGAAAGCGC 240
Qy 81 ThrSerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSer 100
Db 241 TCCACTATAAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCAACTATCATGATGATGATG 300
Qy 101 AsnGlyGlyLeuGlyGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSer 120
Db 301 GTGGGGGACTTTGGAGGACTCTCTCGAATCTGCAAACTTATTTCAGCCAAAGCATACAGT 360
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThr 140
Db 361 GCTGTGTCAGAAATTCATAAACTCTCTGGGAGCAGCAGTGAATGGGCTTACACAACA 420
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAlaAla 160
Db 421 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCTTTTCGCTGCC 480
Qy 161 GlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAlaIle 180
Db 481 CGGAATGAGGACCGACCGCGGAACTCATGTCAGCCAGGACAGGACAGTAAAAATGCAATA 540
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 541 ACAGTCGGAGCTACGGAAACCTCCGCCCAAGCTTTGGTCTTATGCGGACAAATATCAAC 600
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 601 CATGTGGCAGATTTCTTCATCGTGGACCGACAAAGGATGGACGATCAAAACCGGATGTC 660
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 661 ATGGCACCGGAACGTCATCTATCATCAGCAAGATCTTCTCTTGGACCGGATTCCTCCTTC 720
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
```

```
Db 721 TGGCGGAACCATGACAGTAATATGTCATATGCGGTGGAACGTCCATGGCTACACCGATC 780
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 280
Db 781 GTTGTGGAACGTTGGACAGCTTCGTGAGCATTTTGTGAAAAACACAGAGGATCACACCA 840
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaThrAspIleGlyLeuGlyTyr 300
Db 841 AAGCCTTCTCTATTAAAGCGGCACTGATTCGCCGTGCGAGCTGACATCGGCTTGGCTAC 900
Qy 301 ProSerGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 320
Db 901 CCGAACGGTAACCAAGGATGGGAGTGACATTTGGATAAATCCCTGGAAGCTTGGCTAT 960
Qy 321 ValAsnGluThrSerSerLysSerThrAsnGlnLysAlaThrTyrSerPheThrAlaGln 340
Db 961 GTGAACGAGTCCAGTCTCTATCCACAGCCACCAAAAGCGAGCTACTCTGTTTACTGCTACT 1020
Qy 341 SerGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrSerAla 360
Db 1021 GCGGCAAGCCTTTGANAATCTCCCTGTATGGTCTGATGCTCCCTTCGAGCACACTGCT 1080
Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyr 380
Db 1081 TCGGTAAACGTTGTCAATGATCTGGACCTTGTCAATTCGCTCCAAATGGCACACAGTAT 1140
Qy 381 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 400
Db 1141 GTAGGAATAGCTTTACTTCCCATACATGATAACTGGATGGCGGCAATAACGTAGAA 1200
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 420
Db 1201 AATGATTTATTATTAAGCCACCAAGCGGAGCTATACAAATTGAGGTACAGGCTTATAAC 1260
Qy 421 ValProGlnGlyProGlnAlaPheSerLeuAlaIleValAsn 434
Db 1261 GTACCGGTTGGACCAACAGACCTTCTCGTTGGCAATTGTGAAT 1302
```

RESULT 2

US-10-090-624-11
; Sequence 11, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Kunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR FILING DATE: 1999-12-06
; PRIOR FILING DATE: 1999-12-06
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-11

Alignment Scores:
Pred. No.: 3,14e-39
Score: 450.50
Percent Similarity: 47.05%
Best Local Similarity: 31.36%
Query Match: 20.02%
DB: 13
Length: 1977
Matches: 138
Conservative: 69
Mismatch: 170
Indels: 63
Gaps: 15

US-09-985-689A-6 (1-434) x US-10-090-624-11 (1-1977)

Qy 8 ValIyAlaAaspValAlaGlnSerSerTyGlyLeuTyGlyGlnValValAla 27
 Db 433 ATAGGGCCGATACCGTCTGGAACCTCCCTCGGTACGACGGAAGCGGTCTGTTGCC 492
 Qy 28 ValAlaaspThrGlyLeuAaspThrGlyArgAsnAaspSerSerMetHisGluAlaPheArg 47
 Db 493 ATCGTCGATACGGGTATAGACGGGAAC-----CACCCGATCTGAAG 534
 Qy 48 GlyLysIleThrAlaIleTy-----AlaLeuGlyArgThrAsnAlaAsnAaspProAsn 66
 Db 535 GGCAAGGTCATAGCTGTTAGCAGCGCGTCAACGGCGAGTGCAGCCCTCAGATGACGAG 594
 Qy 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyArgGlyThrSerAsnLys----- 84
 Db 595 GGACAGGGAACCCACGTTGCGGTATCGTTGCGGAACCGGCGAGGTTAACTCCCGATPAC 654
 Qy 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMet-----AspSerAsn 101
 Db 655 ATAGGGCTGCGCCCGCGGAGCTCTGTCGGCGTCAAGGTTCTCGGTGCGCAGCGTTCG 714
 Qy 102 GlyGlyLeuGlyGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTySerAla 121
 Db 715 GGAAGCGTCTCCACCATCATCGCGGTGTGTACTGGTCTGCAGCAAGCAAGCAAGTAC 774
 Qy 122 GlyAlaArgIleHisThrAsnSerTpGlyAlaProValAsnGlyAlaTyThrThrAsp 141
 Db 775 GGTATAGGTGTCATCACTCTCCCTCGGCTCTCCAGAGCTCCGAGGACCGACTCC 834
 Qy 142 SerArgAsnValAaspAspTyValArgLysAsnAaspMetAlaValLeuPheAlaIleGly 161
 Db 835 CTCAGTCAGGCGCTCAACACACGCGTGGGCGCGGTATAGTATCTGCGTCCGCGCGC 894
 Qy 162 AsnGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 181
 Db 895 AACAGCGGCGGACACCTACACCGTCTGCGTCAACCGCGCGGCGGAGGTCATAACC 954
 Qy 182 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAaspAsnIleAsnHis 201
 Db 955 GTGCGTGCA-----GTTGACAGCAACGACACA 981
 Qy 202 ValAlaGlnPheSerSerArgGlyProThrLysAaspGlyArgIleLysProAspValMet 221
 Db 982 ATGCGCAGCTTCTCCAGAGGGACCGACCGCGGAGGAGGCTCAAGCGGAGTCTGTC 1041
 Qy 222 AlaProGlyThrPheIleLeuSerAlaAaspSerSerLeuAlaProAaspSerSerPheTrp 241
 Db 1042 GCCCGCGCGTTCATCATAGCCCGCGCGCGCAGC-----GGACACAGCATGGC 1092
 Qy 242 AlaAsnHisAaspSerLysTyAlaTyMetGlyThrSerMetAlaThrProIleVal 261
 Db 1093 ACCCGATAACAGCTACTACACCAAGGCGCTCTGGAACCAAGATGGCCACCGCAGCTT 1152
 Qy 262 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro--- 280
 Db 1153 TCGGCGCTTGGCGCTCATCTCCACAG-----GCCACCGAGTGGACCCCGGAC 1203
 Qy 281 -----LysProSerLeuLeuLysAlaAlaLeuIleAlaIleAlaThrAspIleGly--- 297
 Db 1204 AAGGTGAAGACCGCTCATCGAGCGCGCATAGTCGCCCCCAAGGAGATACCGGAC 1263
 Qy 298 LeuGlyTyProSerGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu--- 316
 Db 1264 ATCCCTTAC-----GGTGGGGTAGGTGAACGTCTACAAGGCCATCAAG 1308
 Qy 317 -----AsnValAlaPheValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThr 334
 Db 1309 TACAGCAGCTACGCCAAGCTCACCTTCACCGGCTCGTCCGCGACCAAGGAGCGCCACC 1368
 Qy 335 TyrSerPheThrAlaGlnSerGlyLysProLysIleSerLeuValTrpSerAspAla 354

US-09-985-689A-6 (1-434) x US-10-090-624-2 (1-1236)

Qy 12 ValAlaGlnSerSerTyGlyLeuTy-----GlyGlnGlyGly 24
 Db 24 GTCTGAGCTCAAGTTATGGCACTTACGTTTGGAACTTGGATATGATGTTCTCGAAT 83
 Qy 24 nValValAlaValAlaAaspThrGlyLeuAaspThrGlyArgAsnAaspSerMetHisGly 44
 Db 84 CACATAGGAATAATTGACATGGANTTGAC-----GCTTCTCATCC 125
 Qy 44 uAlaPheArgGlyLysIleThrAlaIleTyAlaLeuGlyArgThrAsnAsnAlaAsn--- 63
 Db 126 AGATCTCCCAAGGAAGAATA-----ATTGGGTGGGTAGATTTTGTCAATGG 170
 Qy 64 -----AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGly 78
 Db 171 TAGGAGTTATCATACGATGACCATGGACATCGAATCATGCTAGCTTCAATAGCAGCTGG 230
 Qy 78 yAsnGlyThrSerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94

RESULT 3
 US-10-090-624-2
 ; Sequence 2, Application US/10090624
 ; Publication No. US20020132335A1
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, MICO
 ; APPLICANT: SHIMOJO, Tomoko
 ; APPLICANT: ASADA, Kiyozo
 ; APPLICANT: KATO, Ikunoshin
 ; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
 ; FILE REFERENCE: TAKAKURA-6
 ; CURRENT APPLICATION NUMBER: US/10/090,624
 ; PRIOR FILING DATE: 2002-03-06
 ; PRIOR APPLICATION NUMBER: 09/445,472
 ; PRIOR FILING DATE: 1999-12-06
 ; PRIOR APPLICATION NUMBER: 151969/1997
 ; PRIOR FILING DATE: 1997-06-10
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: Patent in version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 1236
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Synthetic
 ; US-10-090-624-2

Alignment Scores:
 Pred. No.: 1,096-34 Length: 1236
 Score: 406.50 Matches: 136
 Percent Similarity: 44.22% Conservative: 63
 Best Local Similarity: 30.22% Mismatches: 169
 Query Match: 18.07% Indels: 82
 DB: 13 Gaps: 17

Db 231 TACTGGACGACGAAGTAATGCGCAAGTACAGGATGGCTCCAGGAGCTAAGCTGGCGG 290
Qy 94 eGlnSerValMet-----AspSerHenglyGlyLeuGlyLeuProSerAsnValse 112
Db 291 AATTAAGGTTCTAGTGGCCGATGGTTCTGGAACATATCTACTATAATTAAGGGAGTTGA 350
Qy 112 rThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 351 GTGGCGCGTGTATACAAAGATAGTACGGAATTAAGGTCATTAACTTTCTCTGGTTC 410
Qy 132 aProValAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyrValArgLysAs 152
Db 411 AAGCCAGAGCTCAGATGCTAGCTCAGCTCTAACTCAGGCTGTGTTAATGCGCGTGGGATCC 470
Qy 152 nAspMetAlaValLeuPheAlaAlaGlyAsnGlyProAsnGlyThrIleSerAl 172
Db 471 TGAATAGTTGTTGGTTCGCGCTGGAACAGTGGACCTAACAGTATACATCGGTTTC 530
Qy 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh 192
Db 531 TCCAGCAGCTGCAAGCAAGTATTATACAGTTGGAGCC----- 567
Qy 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrIly 212
Db 568 -----GTTGACAGTATGATTTATAACAGCTTCTCAAGCAGAGGGCCAACTGC 617
Qy 212 sAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSe 232
Db 618 AGCGGCGAGCTTAAGCCTGAGGTTGTTGCTCAGGAAACTGGATAATTGCTGCCAGAGC 677
Qy 232 rSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGl 252
Db 678 AAGT-----GGAACCTAGCTGGGTCAACCAATTAATGACTATTATACACAGACTCC 728
Qy 252 yGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPh 272
Db 729 TGGACATCAATGCAACTCTCAGTAGCTGTTGATTCGAGCCCTCTGTCCTCAA----- 783
Qy 272 eIleLysAsnArgGlyIleThrPro-----LysProSerIleuLeuLysAlaIle 289
Db 784 ----GCAACCGAGCTGGATCCAGACAAAGTAAACAGCCCTCATAGAACTGCTGA 839
Qy 289 uIleAlaGlyAlaThrAspIleGly---LeuGlyTyrProSerGlyAsnGlnGlyTrpGl 308
Db 840 TATCGTAAAGCCAGATGAATAGCGGATATAGCTTAC-----GGTGCAGG 884
Qy 308 yArgValThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSe 325
Db 885 TAGGTTAATGCATACAGGCTATAAACTACGATACTATGCAAGCTAGTGTCTACTGG 944
Qy 325 rSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAlaGlnSerGlyLysProLe 345
Db 945 ATATGTTGCCAACAAAGCCAGCCAACTCACCAGTTGTTATTAGCGGAGCTTCGTTCTGT 1004
Qy 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrSerAlaSerValThrLeuVa 365
Db 1005 AACTGCCACATATATCTGGGACAATGCCAAAT----- 1035
Qy 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPh 385
Db 1036 -AGGACCTTATCTTACCTCTTACGATCCCAATGGAAACAG---GTTGACTACTCTTA 1091
Qy 385 eThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1092 CACGCGCTTACTAT-----GGATTCGAAAGGTTGGTTATTA 1127
Qy 405 nAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsnValProGlnGlyPr 425
Db 1128 CAACCCAACTGATGGAACATGGCAATTAAGGTTGTAAGCTACAGC-----GGAAG 1178
Qy 425 oGlnAlaPheSerLeuAlaIleValAsn 434

Db 1179 TGCAAACTATCAAGTAGATGTGGTAAGT 1206
RESULT 4
US-10-090-624-15
; Sequence 15, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-15
Alignment Scores:
Pred. No.: 2,17e-34 Length: 1962
Score: 406.50 Matches: 136
Percent Similarity: 44.22% Conservativeness: 63
Best Local Similarity: 30.22% Mismatches: 169
Query Match: 18.07% Indels: 82
DB: 13 Gaps: 17
US-09-985-689A-6 (1-434) x US-10-090-624-15 (1-1962)
Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 420 GTCTGCAGCTCAAGTATGGAACCTTACGTTGGAACTTGGGATATGATGTTCTTGAAT 479
Qy 24 nValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 480 CACATAGGAATTAATGACACTGGAAATGAC-----GCTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAlaAsn-- 63
Db 522 AGATCTCCAAGGAAAGTA-----ATTGGGTGGTAGATTTTGTCAATGG 566
Qy 64 -----AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 567 TAGGAGTTATCCATACGATGACCATGGACATGGAACTCATGTAGCTTCAATAGCAGCTGG 626
Qy 78 yAsnGlyThrSerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATAGCAAGTACAGGAAATGGCTCCAGAGCTAAGCTGGCGG 686
Qy 94 eGlnSerValMet-----AspSerAsnGlyGlyLeuGlyGlyLeuProSerAsnValse 112
Db 687 AATTAAGGTTCTAGTGGCGGATGGTTCGGAAGCATATCTACTATAATTAAGGGAGTTGA 746
Qy 112 rThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGCGCGTGTATACAAAGATAGTACGGAATTAAGGTCATTAACTTTCTCTTGGTTC 806
Qy 132 aProValAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyrValArgLysAs 152
Db 807 AAGCCAGAGCTCAGATGCTAGCTTACGCTCTAAGTCAGGCTGTTAATGAGCGGTGGGATGC 866
Qy 152 nAspMetAlaValLeuPheAlaAlaGlyAsnGlyProAsnGlyThrIleSerAl 172


```
Db 867 TGGAATTAGTTGTTGGTTCGCCCTCGAAACAGTGGACCTAAACAGTATACATCGGTTTC 926
Qy 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh 192
Db 927 TCCAGCAGCTGCGAAGCAAGTATTACAGTTGGAGCC----- 963
Qy 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrIly 212
Db 964 -----GTTGCAAGTATATGTTTATAACAAGCTTCTCAAGCAGAGGGCCAACTGC 1013
Qy 212 sAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSe 232
Db 1014 AGACGCGAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCT 1073
Qy 232 rSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGl 252
Db 1074 AAGT-----GGAACCTAGCAGTGGGTCAACCAATTAATGACTATTACACAGCAGCTCC 1124
Qy 252 yGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPh 272
Db 1125 TGGGATCATCATGCAACTCTCAGTAGCTGTGATGAGCCCTCTTCTCCAA----- 1179
Qy 272 eIleLysAsnArgGlyIleThrPro-----LysProSerLeuLeuLysAlaAlaLe 289
Db 1180 ----GCACACCCGAGCTGGACTCCAGCAAAAGTAAACACAGCCCTCATAGAAACTGCTGA 1235
Qy 289 uIleAlaGlyAlaThrAspIleGly---LeuGlyTyrProSerGlyAsnGlnGlyTrpGl 308
Db 1236 TATCGTAAAGCCAGATGAATAGCCGATATAGCTTAC-----GGTGCAGG 1280
Qy 308 YArgValThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSe 325
Db 1281 TAGGGTTAATGCATACAGAGCTATAAATACGATACTATGCAAGCTAGTGTCACTGG 1340
Qy 325 rSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAlaGlnSerGlyLysProLe 345
Db 1341 ATATGTTGCCAACAAAGGCGACCAACTCACCAGTTCGTTATTAGCGAGCTTCGTTGCT 1400
Qy 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrSerAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATATTACTGGGCAATGCCAAT----- 1431
Qy 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPh 385
Db 1432 -AGCGACCTTGATCTTTACCTCTACGATCCCAATGGAACCCAG---GTTGACTACTCTTA 1487
Qy 385 eThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1488 CACCGCTACTAT-----GGATTCGAAAGTGTGTTATTATTA 1523
Qy 405 nAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsnValProGlnGlyPr 425
Db 1524 CAACCCCACTGATGGAACATGGACAATTAAAGTTGTAAGTACAGC-----GGAAG 1574
Qy 425 oGlnAlaPheSerLeuAlaIleValAsn 434
Db 1575 TGCAAACTATCAAGTAGATGTGTAAGT 1602
```

RESULT 5

```
US-10-156-761-5701
; Sequence 5701, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
```

```
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5701
; LENGTH: 3624
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3624)
; US-10-156-761-5701

Alignment Scores:
Pred. No.: 4,29e-29 Length: 3624
Score: 362.00 Matches: 135
Percent Similarity: 43.88% Conservative: 62
Best Local Similarity: 30.07% Mismatches: 186
Query Match: 16.09% Indels: 66
DB: 14 Gaps: 17

US-09-985-689A-6 (1-434) x US-10-156-761-5701 (1-3624)

Qy 8 ValLysAlaAspValAlaGlnSerSer-----Tyr 17
Db 556 GTCAGGCGGACATGGCGGAGACACCGCAGATCGTAGCGGGCGGCGGGAGCGCC 615
Qy 18 GlyLeuTyrGlyGlnGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 616 GGGCTCAGCGGCGACGGCGTCACCGTCGCGGTCTCGACACCGCGCTCGACAC----- 669
Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGly 57
Db 670 -----ACTACCCCGACCTCGCGCGCGGTGTCGCGAGCAAGAGCTTCATC 717
Qy 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 718 CACGGGAGGAGGTGCGGACCGCAACCGGCAACGGGACCCACGTCACCTCGCGTCGCG 777
Qy 78 GlyAsnGlyThrSerAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 93
Db 778 GGCAGCGCGCGCTCCGACGCGAGCGCGCGCTCGCGCGCGCTCGCGCGCTCGCGCTCGCC 837
Qy 94 PheGlnSerValMetAspSerAsnGlyGlyLeuGlyLeuProSerAsnValSerThr 113
Db 838 GTCGCAAGTGTCT---AGCGACAGGCGCGGAGCGAGTCCAGATCATCGCGGC 894
Qy 114 LeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro 133
Db 895 ATGGAATGGCGCGCGGAGCGTGTGTCAGGATCGTCTCGATGAGCCTCGGATCG--- 951
Qy 134 ValAsnGlyAlaTyrThrThrAsp-----SerArgAsnValAspAspTyrValArgLys 151
Db 952 ACCGAGGCGAGCGGAGCGACCCCATGCGCGAGCGCGTCCGACACCTCTCCGAGGAG 1011
Qy 152 AsnAspMetAlaValLeuPheAlaGlyAsnGluGlyProAsnGlyGlyThrIleSer 171
Db 1012 ACCGCGCGCTCTCTGTCGTCGCGCGGAAACACCGGTGCGC---CCCTCTCGATCGC 1068
Qy 172 AlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSer 191
Db 1069 TCGCGCGCGCGCGCGCTCGCGCTGACCTCGCGCGC----- 1107
Qy 192 PheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThr 211
Db 1108 -----GTCGACTCATCCGCGCGCGCGCTTCTACCGCGCGCGCGCGCGCGC 1155
Qy 212 Lys---AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAla 230
Db 1156 CACGCGGACAAAGCGCTTCAGCGCGCGCGCTCGCGCGCGCGCTCGACATCCGCGCGCC 1215
```

```

Qy 231 ArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyrAlaTyr 250
Db 1216 CGTCCGAGTCCGCCCCGACCGCTAC-----TACACCTCC 1254
Qy 251 MetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGlu 270
Db 1255 ATGAGCGGTACGTCGATGCGAGCGCCCGATGTCGGGGGTCCGCGCTCTCTCGCCGAG 1314
Qy 271 HisPheIleLysAsnArgGlyLeuThrProLysProSerLeuLeuLysAlaAla----- 288
Db 1315 CAGCACCCGACGAGCGCGCGCGCTCAAGGACGCGCTGATGTCACGTCGAGCAAA 1374
Qy 289 LeuIleAlaGlyAlaThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTrpGly 308
Db 1375 CTCGAGCCTCCGCTATATGATGTTGGGGCGGTGCGGTGAGTGTCCGCGACCGCTCGGC 1434
Qy 309 ---ArgValThrLeuAspLysSerLeuAsnValAlaPheValAsnGlnThrSerSerLeu 327
Db 1435 GCCCGGTACCCGCGAGCGCGCGCGCTCCGCTTCCACCGTGGCGCCCGATGAGCC 1494
Qy 328 SerThrAsnGlnLysAlaThrTyrSerPheThrAlaGlnSerGlyLysProLeuLysIle 347
Db 1495 GATCGACCCGCTACAGAAAGCGGTCACTACTCCAACTCTCCGACACGAGCGTCCGAGTTG 1554
Qy 348 SerLeuValTrpSerAspAlaProAlaSerThrSerAlaSerValThrLeuValAsnAsp 367
Db 1555 AGCTTCGCGTGGCGGGCGCGCGCGTGGTGTGCGC-----ACCTCTGCC----- 1599
Qy 368 LeuAspLeuValIleThrAlaPro-----AsnGlyThrLysTyrValGlyAsn 383
Db 1600 ---GACACGCGACTCACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAC 1656
Qy 384 AspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPhe 403
Db 1657 GGTCCAGGCTCCGCTG-----GGAGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1698
Qy 404 IleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsnValProGln 423
Db 1699 -----GCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1749
Qy 424 GlyProGlnAlaPheSerLeuAlaIle 432
Db 1750 GAGCGCGAGCGGTACTCCCTCACCGGTG 1776

```

RESULT 6

US-10-156-761-1

; Sequence 1, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156, 761

; CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 1

; LENGTH: 9025608

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (4187715)

; OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1

Alignment Scores:

Pred. No.:	4,86e-24	Length:	9025608
Score:	362.00	Matches:	135
Percent Similarity:	43.88%	Conservative:	62
Best Local Similarity:	20.07%	Mismatches:	186
Query Match:	16.09%	Indels:	66
DB:	14	Gaps:	17

US-09-985-689A-6 (1-434) x US-10-156-761-1 (1-9025608)

```

Qy 8 ValLysAlaAspValAlaGlnSerSer-----Tyr 17
Db 6918813 GTCGAGGCGGACATGTCGCGAGGACGACGCGCAGATCGGTACGCGGCGCGGTGGAGGCC 6918872
Qy 18 GlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 6918873 GGGCTCAGCGGCGAGCGGTGTCACCGTCCGCGTGTGTCACACCGCGCGCTCGACATCC- 6918926
Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGly 57
Db 6918927 -----ACTCACCCCGACCTCGCGCGCGGTGTCCCGAGCAAGAGCTTCATC 6918974
Qy 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 6918975 GACGCGGAGGAGTTCGCGACCGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919034
Qy 78 GlyAsnGlyThrSerAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 93
Db 6919035 GCACGCGGCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919094
Qy 94 PheGlnSerValMetAspSerAsnGlyGlyLeuGlyGlyLeuProSerAsnValSerThr 113
Db 6919095 GTCGCAAGTGCTC---AGCGACGCGCGCGCGGAGCGAGTCCAGATCATCGCGCGCG 6919151
Qy 114 LeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro 133
Db 6919152 ATGGAATGGCGCGCGCGGACGTGCGTCCAGGATCGTCTCGATGAGCCTCGGATCG- 6919208
Qy 134 ValAsnGlyAlaTyrThrAsp-----SerArgAsnValAspAspTyrValArgLys 151
Db 6919209 ACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919268
Qy 152 AsnAspMetAlaValLeuPheAlaGlyAsnGlyProAsnGlyGlyThrIleSer 171
Db 6919269 ACCGCGCGCTCTTCGTCGTCGCGCGGAAACACCGGTGCC---CCCTCTCGATCGCG 6919325
Qy 172 AlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSer 191
Db 6919326 TCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919364
Qy 192 PheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThr 211
Db 6919365 -----GTGACTCATCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919412
Qy 212 Lys---AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAla 230
Db 6919413 CAGCGCGCAACGCCCTCAAGCCGACCTCCGCCACCCCGCGCTCGACATCCCGCGCGCGCG 6919472
Qy 231 ArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyrAlaTyr 250
Db 6919473 CGCTCCAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919511
Qy 251 MetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGlu 270
Db 6919512 ATGAGCGGTACGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919571
Qy 271 HisPheIleLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAla----- 288
Db 6919572 CAGCACCCGAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919631

```



```

Db      1696  GTC-ACGCTGAACTCGCGTCAGCGCGACCAACCCCAAGGCGCTCGCGCTCGGTCTCGTCTCGG 1754

QY      379  -----LysTyValcIlyAsnAspPheThrAlaProTyRAspAsnAsnTrpAspGlyAr 396
Db      1755  CTTCTTCAGACTCGGCGCGACGAAGGTACAGGTCTCCGGCG-----GGCGGCAA 1802
QY      396  gaAsnValGlu-----AsnValPheIleIenAlaProGlnSerGlyThrTyTh 413
Db      1803  GGCTCGGTCTCACTTCACCGTCAACAGGAAGCTGGCGCGCACCAACGGACGGCGGTACTC 1862
QY      413  rValGluValGlnAla 418
Db      1863  GGCGTACGTACGGCC 1878

RESULT 8
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHISA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 2,52e-23 Length: 9025608
Score: 355.50 Matches: 138
Percent Similarity: 43.90% Conservative: 49
Best Local Similarity: 32.19% Mismatches: 167
Query Match: 15.80% Indels: 73
DB: 14 Gaps: 17

US-09-985-689A-6 (1-434) x US-10-156-761-1 (1-9025608)
QY      18  GlyLeuTyRgIyGInGInValValAlaAlaAspThrGlyLeuAspThrGlyArG 37
Db      4132505  GGGTACGACGCGCAAGGCGGTCAAGATCGCGTCTGTGACACCGGTGCGAC----- 4132455
QY      38  AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaIleTyRAlaLeuGly 57
Db      4132454  -----GGACCCACCCGGACCTCAAGACACAGGTGGCCGAGTCCAAGACTTCTCC 4132404
QY      58  ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValIaGlySerValLeu 77
Db      4132403  GCGCGCGCGCGACGCGCGCACCACTTCGGTCACGCGACGACGTCGCGTCCATCGCGGCG 4132344
QY      78  GlyAsnGlyThrSerAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 93
Db      4132343  GGCACCGCGCGCAAGTCCACGGCAAGTACAAGGGTGTGGCGCGCGCGACGATCCTC 4132284
QY      94  PheGlnSerValMetAspSerAsnGlyLeuGlyLeuProSerAsnValSerThr 113

```



```
US-09-974-300-1934
Alignment Scores:
Pred. No.: 4,81e-27 Length: 1329
Score: 337.50 Matches: 108
Percent Similarity: 45.18% Conservative: 42
Best Local Similarity: 32.53% Mismatches: 115
Query Match: 15.00% Indels: 67
DB: 9 Gaps: 13

US-09-985-689A-6 (1-434) x US-09-974-300-1934 (1-1329)
QY 11 AspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyValValAlaValAlaAsp 30
DB 406 GAATGTGTGAGAAACATCAGACCTGACAGGCAAGGAGTGACAGTGGTGTGATGAT 465
QY 31 ThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIle 50
DB 466 ACGGGCGTA-----TACCCTCAGCAAGATCTTTGAAGCAGGATC 504
QY 51 ThrAlaIleTyr---AlaLeuGlyArgThrAsnAlaAsnAspProAsnGlyHisGly 69
DB 505 AGGCTTTTCAAGACTTTATCAACAGAGAAACAGACCTATGATGACATGGCAGGC 564
QY 70 ThrHisValAlaGlySerValLeuGlyAsnGlyThrSerAsn-----LysGly 85
DB 565 ACACATCGCGCGGTGATGCTTGGGAAACGGAGCGGCTCATCGGTCAGTACCGCGGA 624
QY 86 MetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSerAsnGlyGlyLeuGly 105
DB 625 CCTGCTCTTGAGCAGAACTTGTGCGTGTAAAGTATTTGGACAAATGGGATCC---GGA 681
QY 106 GlyLeu-----Pro 108
DB 682 TCGCTCGAAACCGTCATTCAAGCGGTAGATTGGTGCTCAATTCATTAAGGAAATCCT 741
QY 109 SerAsnValSerThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsn 128
DB 742 GATGATCCGATCGACATTATTCAATGATCTGGTGGCAGAGCCTTGGCGTACGAGAT 801
QY 129 SerTrpGlyAlaPro-----ValAsnGlyAlaTyrThrThrAspSerArgAsn 144
DB 802 GAAGAAGAGATCCAGTCGTAAAGCTGTTCAATGCGAGCATGGGACGCA-----849
QY 145 ValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAlaAlaGlyAsnGlyGly 164
DB 850 -----GGCATTGTTGTATGTCGGCAGCGCGCAACTCCCGT 895
QY 165 ProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 184
DB 886 CTGTATGCCCAACGATTCGCCAGCCGGGTGTCAGCAGCAAGATTATTACAGTCGGAGC 945
QY 185 ThrGluAsnLeuArgSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGln 204
DB 946 TTGGATGAC---AGGGATACAGTCAGCCGGGAGGATGACGAT-----GTGCGCTCT 993
QY 205 PheSerSerArgGlyProThrLysAspGlyArgIleLysProAspValMetAlaProGly 224
DB 994 TATTCAGCAGAGGCGGCAATCTATGCTCAAGTCRAACCGGACTTCTGCTGACCGGCG 1053
QY 225 ThrPheIleLeuSerAlaArgSer-----SerLeuAlaProAspSerSerPheTrpAla 242
DB 1054 ACAATATTACGTCGCTTCGTTCCACCGGATCTTTCTCGATAGCTGCAAAACAAAC 1113
QY 243 AsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAla 262
DB 1114 AGAGTCGGCAGCAAAATATATGACATTTGTCGGAACCTCGATGGTACGCCGATCTGCCA 1173
QY 263 GlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysPro 282
DB 1174 GGAATTGGCGCA-----CTTATCTTCAGCAAGCCCGGCGCACAGAACT 1218
QY 283 SerLeuLeuLysAlaAlaLeuIleAlaGlyAlaThrAspIle-----296
```

```
DB 1219 GATGAGTCAACAGCTGCTTAATGACGGT---ACCGATTATGGAAGATCGGATCCA 1275
QY 297 -----GlyLeuGlyTyrProSerGlyAsnGln 305
DB 1276 AATGTTTACGGTGCAGGATACATCAACGCAAGACAA 1311
RESULT 11
US-10-090-624-5
; Sequence 5, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hi-karu
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyo-ozo
; APPLICANT: KATO, Ikunobu
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 5
; LENGTH: 4765
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-5
Alignment Scores:
Pred. No.: 5,82e-24 Length: 4765
Score: 317.00 Matches: 131
Percent Similarity: 39.8% Conservative: 50
Best Local Similarity: 28.85% Mismatches: 142
Query Match: 14.09% Indels: 131
DB: 13 Gaps: 15

US-09-985-689A-6 (1-434) x US-10-090-624-5 (1-4765)
QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 1024 GGCAATGGTATGACATTGCTATGATGCTGCTGCTTTTAGCTACTACTACGCT 1083
QY 41 -----SerMetHisGluAlaPheArgGly 48
DB 1084 GTTCCACTTGGCCAGTACAACTTACTATGATGCTGCTGCTTTTAGCTACTACTACGCT 1143
QY 49 LysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAlaAsnAspProAsn-----66
DB 1144 CCTCTCAACTACGTGCTTGA-----GAATAGATCTTACCGAGAA 1185
QY 67 -----GlyHisGlyThrHisValAlaGlySerValLeu 77
DB 1186 TATGCAATATTGGTGGGATGCTACGGTCAAGGAACTCAGTAGCTGGAAGTGTGCT 1245
QY 78 GlyAsnGlyThrSerAsn-----83
DB 1246 GGTTCAGACAGCAACAATGATGCTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1305
QY 84 -----LysGly 85
DB 1306 GTGTTCTCAAGACTTATGTTGGGATTTACGAGCTTACCAAGCTTACCAAGCTGTCAGGCT 1365
QY 86 MetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSerAsnGlyGlyLeuGly 105
DB 1366 GTTCTCCAGTGGCCCAATATAGGCAATAGAGTCTTTAGGAGTGTATGGA-----CGG 1419
```

QY 106 GlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle 125
Db 1420 GGTAGCATGGGATATTATAGAGGTATGACATACGCAACCCATGGTGAGACGTT 1479
QY 126 HisThrAsnSerTyrGly-----AlaProValAsnGlyAlaTyrThrAspSerArg 143
Db 1480 ATAAGCATGAGTCTCGTGGGAATGCTCCATATTAGATGGTACTGATCCAGAAAGCGTT 1539
QY 144 AsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAlaAlaGlyAsnGlu 163
Db 1540 GCTGTGATGAGCTTACGAAAGTACGGTGTGTTATTCGTAATAGTCGAGGAATGAA 1599
QY 164 GlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 183
Db 1600 GGTCTCGCATTAACATCGTGTGGAGTCTCGTGTGTCACAAAGGCAATACCTGTGGA 1659
QY 184 AlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAsp----- 197
Db 1660 GCTGCTGCA---GTGCCCATTAACGTGGAGTTATGTTCCCAAGCACTGGATATCCT 1716
QY 198 -----AsnIleAsnHisValAlaGlnPhe 205
Db 1717 GATTACTATGATTCATTACTTCCCGCCTACACAAACGTT---AGATAGCATTTCTTC 1773
QY 206 SerSerArgGlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThr 225
Db 1774 TCAAGCAGAGCGCGAGATAGATGTGAATAAACCAATGATAGTGGTCCAGGTAC 1833
QY 226 PheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAsp 245
Db 1834 GGAATTTACTCATCCCTCCCGCATGTGATTGGCGAGCTGACTTC----- 1878
QY 246 SerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnVal 265
Db 1879 -----ATGCTGGNACTTCGATGGCTACTCCATGCTCAGCGGTGCGTT 1923
QY 266 AlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSerLeuLeu 285
Db 1924 GCATCTCTCATAGCGGG---GCAAGCGCGAGGAATATACTACATCCAGATATAAAT 1980
QY 286 LysAlaAlaLeuIleAlaGlyAlaThrAspIle-----GlyLeuGly 299
Db 1981 AAGAGGTTCTTGAGCGGTGCAACTGCTGAGGAGATCCATATATCTGGCGAGAAG 2040
QY 300 TyrProSerGlyAsnGlnTyrTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 319
Db 2041 TACACTGAGCTTGACCAAGGTATGTTCTGTTAACTTACCAAGTCTCGGAAATC--- 2097
QY 320 PheValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAla 339
Db 2098 -----CTTAAGGCT 2106
QY 340 GlnSerGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrSe 359
Db 2107 ATAAACGGCACCATCTCCCAATTTGATCACTGGCGACAGCAAGTCTACAGGACTTT 2166
QY 359 rAlaSerValThrLeuValAsnAspLeu-----AspLeuValIle 372
Db 2167 CGCGAGT---ACTTGGGTGTGGAGCTTATAAGAGGTCTCTACGCAAGGAATCTATACCT 2223
QY 372 eThrAlaProAsnGlyThr--LysTyrValGlyAsn 383
Db 2224 GACATTGTGAGTGGCATTAAAGTACGTAGGGGAC 2259

RESULT 12

US-10-084-846A-113
; Sequence 113, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLERWEG, AGNES
; APPLICANT: TREFFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS

; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084.846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 113
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: avx16 dna: partial sequence of coding strand 1; nucleotide 1
; OTHER INFORMATION: corresponds to nucleotide 56,198 of coding strand 1.
US-10-084-846A-113

Alignment Scores:

Pred. No.:	1-55e-22	Length:	1560
Score:	297.50	Matches:	120
Percent Similarity:	41.36%	Conservative:	50
Best Local Similarity:	29.20%	Mismatches:	166
Query Match:	13.22%	Indels:	75
DB:	15	Gaps:	19

US-09-985-689A-6 (1-434) x US-10-084-846A-113 (1-1560)

QY	21	GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer	40
Db	451	GGGCAGGAGGTGACGGCGTACGTATCATGCACACCGCGCTC-----CGC	492
QY	41	SerMetHisGluAlaPheArgGlyLysIleThrAlaIleTyr---AlaLeuGlyArgThr	59
Db	493	ATCACCCACAGGACTTCGGCGCGCGCTCTTACGGCTACGACGCCATCGACACACGAC	552
QY	60	AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn	79
Db	553	AAACCGGCCAGGACGGCCACGCGCACGCGCAGCGTGGCGGCACGCTCCCGGCAAC	612
QY	80	GlyThrSerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAsp	99
Db	613	GCC-----TACGGCGTCCCAAGAGCGCCAGATCTGATGGGCTCCGCGTCTGCTG	660
QY	100	SerAsnGlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyr	119
Db	661	AAACAACCTCGGCGCAGGCGCACCCAGGTCGTCGCGCGCATCGACTGGTCCGCCGG	720
QY	120	SerAlaGlyAlaArgIleHisThrAsn---SerTrpGlyAlaProValAsnGlyValTyr	138
Db	721	AACGCGCTCAAGCCGCGCGCTCGCCCAACATGCTCTCGCGCGCGCGCGCACGCGCTC	780
QY	139	ThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPhe	158
Db	781	GACACGGCGGTACGCAAC-----GCCATGGCTCCGGCGTCACTTCGCGGTG	828
QY	159	AlaAlaGlyAsnGlyGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsn	178
Db	829	GGCGCGCGCAACGAGTCGACCAACCGCTCCACAGGTCAC---CCGCGCGCTCCCGAG	885
QY	179	AlaIleThrValGlyValAlaThr-----GluAsnLeuArgProSerPheGlySerTyrAla	196
Db	886	GCATTCACGGTCCGCGCGCGACCGACAGTCGACGCGCAAGCGCGGTACTTCCACTACGC	945
QY	197	AspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIle	216
Db	946	TCCGTCCTC-----	954
QY	217	LysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPro	236
Db	955	-----GACCTCTTCCGCGCGCGGTCTTCCTCATCACTCGGCC-----	990

QY 370 LeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyr 389
Db 57563 GTCCAGCTGATCCGCCCGGAGCGGACGCGGTACACGCTCAAGTCGTAC----- 57610
QY 390 AspAsnThrAspGlyArgAsnValGluAsnValPhe---IleAsnAlaProGln 408
Db 57611 ---GGCACCAGCGGCGAGTTCGGGACACATCAACACCGTACTCGGTGAACCGCTCCCTCG 57667
QY 409 -----SerGlyThrTyrThrValGluVal 416
Db 57668 GAGCGCGCCACGCGACGTCGAACTGCGGGTG 57700

RESULT 14

US-10-084-846A-2/c
; Sequence 2, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLER, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent in Ver. 3.2
; SEQ ID NO 2
; LENGTH: 59816
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-2

Alignment Scores:

Pred. No.: 3 52e-20 Length: 59816
Score: 297.50 Matches: 120
Percent Similarity: 41.36% Conservative: 50
Best Local Similarity: 29.20% Mismatches: 166
Query Match: 13.22% Indels: 75
DB: 15 Gaps: 19

US-09-985-689A-6 (1-434) x US-10-084-846A-2 (1-59816)

QY 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 3169 GGGCAGGGAGTACGGGTACGTATCATCGACACCGGGTCTC-----CGC 3128
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaIleTyr---AlaLeuGlyArgThr 59
Db 3127 ATCACCACAGCGATTCGGGCGCGGCTCTCTACGGCTACGACCGCATCGACACGAC 3068
QY 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 3067 AACACCCCGCAGGACCGCCACCGCCACGCGACGACGTCGGCGCGGACGTCGCGGCAAC 3008
QY 80 GlyThrSerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAsp 99
Db 3007 GCC-----TACGGCGTCGCCAAGAGCGCAAGTCTAGCGTCCCGTCTGCTG--- 2960
QY 100 SerAsnGlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyr 119
Db 2959 AACAACTCCGGCCAGGGACACCGCCCGAGTCTGTCGGCGCATCGACTGGTGGCGCGG 2900
QY 120 SerNlaGlyAlaArgIleHisThrAsn---SerTrpGlyAlaProValAsnGlyAlaTyr 138
Db 2899 AACGCCGTGAAGCGCGCGTGGCCAAACATGTCTCTCGCGCGCGCGCGACGCGGCTCT 2840
QY 139 ThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetAlaValLeuPhe 158

Db 2839 GACACGCGCTACGCAAC-----GCCATGCGCTCCGCGGTCACCTTCGCGGTG 2792
QY 159 AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsn 178
Db 2791 GCGGCGCGCAACGAGTCGACCAACGCTCCACGAGTCA---CCGCGACGCTCACCGAG 2735
QY 179 AlalleThrValGlyAlaThr-----GluAsnLeuArgProSerPheGlySerTyrAla 196
Db 2734 GCCATCAGGTGCGCGCGACGACCGAGCTCGAGCCCAAGCCGCTACTCCAACTACGCG 2675
QY 197 AspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIle 216
Db 2674 TCCGTCTC----- 2666
QY 217 LysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPro 236
Db 2665 -----GACCTTTCGCGCGCGGTCTGTCATCACTCCGCGC----- 2630
QY 237 AspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMet 256
Db 2629 -----TGAACCTCAAGCGACTCGCGACCAACCACTCTCCGTTACGTCGATG 2582
QY 257 AlaThrProIleValAlaGlyAsnValAlaGlnLeuArgLuhisPheIleLysAsnArg 276
Db 2581 GCGACCCCGCACGTCGCGGCGCGCGCTC-----CACCTCGCGCGCAACCCC 2531
QY 277 GlyIleThrProLys-----ProSerLeuLysAlaAlaLeuIleAlaGlyAla 293
Db 2530 TCGGCCACCCCGTCCCGAGTCCGACGCGGTGACGTCGCGCGCACACCGCGGTGTC 2471
QY 294 ThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTrpGlyArgValThrLeuAsp 313
Db 2470 ACCAACCCCGCACGCGGTGCGCCCAACCGGCTCTGTACGTCGCGCGG---CGGACGAC 2414
QY 314 LysSerLeuAsnValAlaPheValAsnGluThrSerSerLeuSerThrAsnGlnLysAla 333
Db 2413 CACCTTCGCGCGCGCGCTTCGAGAAC---ACCGGTGACTACGATCAGTACGACCACTCC 2357
QY 334 ThrTyrSerPheThrAlaGlnSerGlyLysProLysIleSerLeuValTrpSerAsp 353
Db 2356 ACGGTGCGAG-----TCCCGGTGACGCTCTCCGCGTCTCCGCGCAAC 2315
QY 354 AlaProAlaSerThrSerAlaSerVal-----ThrLeuValAsnAspLeuAsp 369
Db 2314 GCGGCTCGGCGCTCGCGTAGAGTCCACATCGTCACACGATCATCGGCGACCTCCAG 2255
QY 370 LeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyr 389
Db 2254 GTCCAGCTGATCGCCCGCGCGACGCGGTACGCTCAGTCAAGTCGTAC----- 2207
QY 390 AspAsnAsnTrpAspGlyArgAsnValGluAsnValPhe---IleAsnAlaProGln 408
Db 2206 ---GGCACCAGCGCGAGTTCGACACATCAACACCGAGTACTCGGTGACGCTCTCTCG 2150
QY 409 -----SerGlyThrTyrThrValGluVal 416
Db 2149 GAGCGCGCAACGCGACGTCGGAACCTCGGGTG 2117

RESULT 15

US-09-927-827-33
; Sequence 33, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramseier, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927, 827
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69

```
; SEQ ID NO 33
; LENGTH: 3788
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(2788)
US-09-927-827-33

Alignment Scores:
Pred. No.: 1,57e-20 Length: 3788
Score: 284.50 Matches: 114
Percent Similarity: 40.76% Conservative: 47
Best Local Similarity: 28.86% Mismatches: 112
Query Match: 12.64% Indels: 122
DB: 10 Gaps: 18

US-09-985-689a-6 (1-434) x US-09-927-827-33 (1-3788)
QY 3 ValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrglyLeuTyrglyGln 22
DB 1739 GTCTGGGCAACGTGCTGATGTCAGTCAGCAGCGCGGATATCGGCACGTGGGCAG 1798
QY 23 GlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMet 42
DB 1799 CAGATCGGCTGGCGGCTCTGGATACCGCAT-----GCGCGCGCG 1840
QY 43 HisGluAlaPhe-----ArgGlyLysIleThrAlaIleTyr----- 54
DB 1841 CATCGCATTTCTTCGTCCGCGGAGCGGCAATGTGGGCGCAATGGGATGTCAGC 1900
QY 55 -----AlaLeuGlyArgThrAsnAlaAsnAsp 64
DB 1901 CGGCGCGGCGCCCAAGCGGTGACGCGGCGGATGCGGCGTGTTCGCGGCGCGCGTG 2020
QY 65 ProAsnGlyHisGlyThrHisValAlaGlySerValLeuGly----- 78
DB 1961 GGGCAGCGGCATGCGACGACATCGCGCGCATCATCGCGCGTTCACCGTTCACCGACTCGAC 1960
QY 79 -----AsnGlyThrSerAsnLys-----GlyMetAlaProGlnAla 90
DB 2021 CTGTATGCCAGTGGGAAGCGGCGGCACTGTGGGAATTTGCGGATGCGCGCGGAGACG 2080
QY 91 AsnLeuValPheGlnSerValMetAspSerAsnGlyLeuGlyGlyLeuProSer--- 109
DB 2081 CAACTCTATGGCTTCAAGTGTCTGACGACGCGCGC-----AACGCGCGCGATTTCGTGG 2134
QY 110 -----AsnValSerThrLeuPheSerGlnAla----- 118
DB 2135 ATGATCAAGCGAGTGCAGCAGTGGCGGATCTCAACGAGCGTGGCGGAGCTGGTGATC 2194
QY 119 -----TyrSerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProVal 134
DB 2195 CACGCGCTCAATCTCAGCTGGCGGCTACTTCGATCGAAAGCTACGGCTGTGGCTTC 2254
QY 135 AsnGlyAlaTyrrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMet 154
DB 2255 -----ACGCGGTTATGCAACAGAGTTGCGCGGCTTATGCGCGGCGCGCGTA 2299
QY 155 AlaValLeuPheAlaAlaGlyAsnGluGly-----ProAsnGlyGly 168
DB 2300 CTGGTAGTGGTGGCGCGGCAACGAAGCGCTGGCTGGCTGATGCGAAGCGCGCGC 2359
QY 169 Thr-----IleSerAlaProGlyThrAlaLysAsnAlaIle 180
DB 2360 ACCTATCCGCGCAACATGGATCTGTGATCAGCATCCGGGCAATCTGGAGACCGCATC 2419
QY 181 ThrValGlyValaThrGluAsnLeuArgPro---SerPheGlySerTyrAlaAspAsnIle 199
DB 2420 GTGGTGGATCGGTGACAGACGACCGCGCACAAATACGCG----- 2461
QY 200 AsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAsp 219
```

Search completed: April 5, 2004, 02:55:48
Job time: 7399 secs

```
DB 2462 -----GTGTGCTATTTTTCATCCGCGGCGGCGGACCGCGATGGCGCTCCAAACCTGAT 2515
QY 220 ValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSer 239
DB 2516 GTGGTGGCGCGCGGCGAAAGATCTGTCCGCTTACTACGGCTTCGACCCG----- 2566
QY 240 PheTrpAlaAsnHisAspSerLysTyrrAlaTyrrMetGlyGlyThrSerMetAlaThrPro 259
DB 2567 -----CGCGACCGCGTCCAGCCTGATGTCGAGATGAGCGCACAGCATGCGCGCACCG 2620
QY 260 IleValAlaGly----- 263
DB 2621 CATGTGTCGCGGTGCTGGCGGGTTTATCCGACGCGCGAGTTTCATCGGCTTTCCG 2680
QY 264 ---AsnValAlaGlnLeu-----ArgGluHisPheIle 273
DB 2681 GACCGGTCAGCAACTGCTGCTGACACCTGCGACCGACCTGCGAGCGGATCGTTCGTG 2740
QY 274 LysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla 293
DB 2741 CAGGCGCAGGGGGTG-----CCGAATTTGATCGG-----ATGCTTGGAGAG 2782
QY 294 ThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTrp 307
DB 2783 ACGTGATTGCGGAGTCCGGATTTGGGATTCGCAACGCGGTTGG 2825
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:40:05 ; Search time 2279.84 Seconds
(without alignments)
5684.703 Million cell updates/sec

Title: US-09-985-689A-6
Perfect score: 2250
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAYNVQGFQAFSLAIWN 434

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=xlh
-Q=/cgn2_1/USPTO.spool/US09985689/runat_31032004.161807.4169/app query.fasta.1.3498
-DB=EST -OPMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bites -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPMT=ptc -NORM=est -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689@cgn2_1.1708@runat_31032004.161807.4168 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -JONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gssl:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	239	10.6	1605	13	BQ622771 CC Contig
2	236	10.5	640	12	BJ395336 BJ395336
3	227.5	10.1	4198	11	AK029048 Mus muscu
4	213.5	9.5	532	29	TA319GL0P
5	213.5	9.5	594	12	BQ393752 B393752
6	207.5	9.2	2141	13	BQ142519 Contig6 M
7	202.5	9.0	601	12	BQ387574 BJ387574
8	200	8.9	508	28	AQ652212 Sheared D
9	200	8.9	771	14	CA320325 UI-M-FW0-
10	198	8.8	1002	29	CNS06D68
11	190.5	8.5	574	29	TA315H10P
12	189	8.4	633	12	BJ369190
13	189	8.4	718	12	BI750157
14	186	8.3	716	28	BZ893395
15	185	8.2	665	13	BQ770462 UI-M-FIO-
16	183.5	8.2	681	14	CB690041 CBST-54-B
17	182	8.1	675	14	CF727824 UI-M-HB0-
18	182	8.1	4662	11	BC060627 Mus muscu
19	181	8.0	530	29	CNS010FO
20	179.5	8.0	545	13	BU575479 TGESTzyb8
21	178.5	7.9	614	9	AJ273402
22	178	7.9	576	14	CD295943
23	177.5	7.9	641	12	BJ393925 BJ393925
24	177	7.9	3091	11	BC011275 Mus muscu
25	174	7.7	594	14	CF846026 pABH036XA
26	174	7.7	2121	28	BZ424995
27	173.5	7.7	650	9	AJ274038
28	173.5	7.7	1029	29	CNS071DW
29	173	7.7	895	13	BQ216158
30	171.5	7.6	616	14	CF138007
31	171.5	7.6	712	14	CD311344
32	171	7.6	564	28	AQ651427
33	170.5	7.6	573	14	CA937626
34	169.5	7.5	655	28	BZ349201
35	168.5	7.5	580	9	AJ273745
36	168.5	7.5	583	9	AJ273947
37	168.5	7.5	593	9	AJ273918
38	168.5	7.5	601	9	AJ273921
39	167.5	7.4	1572	11	AY107161
40	166.5	7.4	601	9	AJ273050
41	166.5	7.4	601	12	BJ365857
42	166.5	7.4	930	13	EX390734
43	166.5	7.4	937	12	BG246418
44	166	7.4	449	12	BJ359939
45	165.5	7.4	604	9	AJ273185

ALIGNMENTS

RESULT 1
BQ622771
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE

BQ622771 1605 bp mRNA linear EST 01-JUL-2002
CC Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
CDNA, mRNA sequence.
BQ622771
BQ622771.1 GI:21649940
EST.
Conidiobolus coronatus (Delacroixia coronata)
Conidiobolus coronatus
Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
1 (bases 1 to 1605)

AUTHORS Freimoser, F.M., Screen, S., Hu, G. and St. Leger, R.J.
 TITLE EST analysis of genes expressed by the zygomycete pathogen
 Conidiobolus coronatus during optimized secretion of proteins
 JOURNAL Unpublished (2002)
 COMMENT Contact: Freimoser F. M.
 Department of Entomology
 University of Maryland
 4112 Plant Sciences Building, College Park, MD 20742, USA
 Tel: 301 405 16 13
 Fax: 301 314 92 90
 Email: ff34@umd.edu.

FEATURES
 source
 1..1605
 /organism="Conidiobolus coronatus"
 /mol_type="mRNA"
 /strain="ARSEF 512"
 /db_xref="taxon:34488"
 /clone_lib="Conidiobolus coronatus ARSEF 512"
 /note="Vector: Unizap; Conidiobolus coronatus was grown in minimal medium supplemented with Manduca sexta cuticle and peptone for 18 hours. A cDNA library was constructed in the unidirectional Lambda vector Unizap."

ORIGIN
 Alignment Scores:
 Pred. No.: 2,46e-13 Length: 1605
 Score: 239.00 Matches: 85
 Percent Similarity: 43.99% Conservative: 43
 Best Local Similarity: 29.21% Mismatches: 83
 Query Match: 10.62% Indels: 80
 DB: 13 Gaps: 14

US-09-985-689A-6 (1-434) x BQ622771 (1-1605)

QY 6 GlylleValValAlaAspValAlaGlnSerSer-----TyrGlyLeu----- 19
 Db 718 GGTGTGTAGAGGTATTGCCGTCCACCAATGCTCTGGGGTCTTGGTGGT 777
 QY 20 -----TyrGlyGlnGly 23
 Db 778 CAACGTGCTAAGCTCGGATCTGCTCTTACTTACACCATGCTGATGCTGCAAGGT 837
 QY 24 GlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHis 43
 Db 838 GTCACTGTTTTCGTTTGTAGTACTGCTGTCATGTCATGTCACCCACATGAC----- 885
 QY 44 GluAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsnAla--- 62
 Db 886 -----TTGGGTGGTGGTGGTGGT-----TGGGGTACTTAACACTGCTGGT 924
 QY 63 -----AsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
 Db 925 GGTAGCAACACTGATGCTCAGGTACCGGTACCGGTACTCTACTGCTGGTACTGCTGGT--- 981
 QY 80 GlyThrSerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMet--- 98
 Db 982 -----ACCACCTATGTTGTTCCCAAGAGGCTAACATTTGCCGTTAAGTCTTAGT 1035
 QY 99 ---AspSerAsnGlyLeuGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGln 117
 Db 1036 GATGATGGTCCGGATCACTACTGGAATTACTCCGTTATTCCTGGTATTGACTGGTGTAG--- 1092
 QY 118 AlaTyrSerAlaGlyAlaAlaArgIleHisThrAsnSerTrpGlyValaProValAsnGlyAla 137
 Db 1093 ---CACTCTGCTCCCAAGAAAGTATTCTCTATGAGTTTAGGAGGTGGTAAGAACATGCT 1149
 QY 138 TyrThrThrAspSerArgAsnValaAspAspTyrValArgLysAsnAspMetAlaValLeu 157
 Db 1150 CTTAACACT-----GCTGTATAACACGCTGTCCAGAGGGAGTTGCTACTGTTGTC 1200
 QY 158 PheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys 177
 Db 1201 ---GCTGTGGTACGATACAGAGATGCTTGTGGTACTCT---CCCGTCTCTCTCT 1254

QY 178 AsnAlaIleThrValGlyValAlaThrGly-----AsnLeuArgProSerPheGlySerTyr 195
 Db 1255 TCCGCCATTACCGTGGTGGCTGACCTGATGATCAATAAAGGGCTTCTTCTTCACTTC 1314
 QY 196 AlaAspAniLeuAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArg 215
 Db 1315 GGTAGCTGTGC----- 1326
 QY 216 IleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaAArgSerSerLeuAla 235
 Db 1327 -----GATATCTTAGCTCTCTGGTGTCAACATTCTCTCCACC----- 1362
 QY 236 ProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSer 255
 Db 1363 -----TGGAGGGATCTTAACACTGCCACCAACACCATCTCTGTTACCTCT 1407
 QY 256 MetAlaThrProIleValAlaGlyAsnValAla 266
 Db 1408 ATGGCTTGGCTTCCCTCATTGCTGGTTAGCTGCT 1440

RESULT 2
 BQ622771
 LOCUS 640 bp mRNA linear EST 08-MAR-2002
 DEFINITION Dictyostelium discoideum cDNA library, SF Dictyostelium
 accession BQ622771.1 GI:19306422
 VERSION BQ622771.1
 KEYWORDS Dictyostelium discoideum
 SOURCE Dictyostelium discoideum
 ORGANISM Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 REFERENCE 1 (bases 1 to 640)
 AUTHORS Urushinara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
 TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadao Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-553-81-6856
 Fax: 81-553-81-6855
 Email: tshini@genes.nig.ac.jp.

FEATURES
 source
 1..640
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="dds38b16"
 /sex="mat A"
 /dev_stage="Slug stage"
 /clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.03e-13 Length: 640
 Score: 236.00 Matches: 69
 Percent Similarity: 49.77% Conservative: 37
 Best Local Similarity: 32.39% Mismatches: 69
 Query Match: 10.49% Indels: 38
 DB: 12 Gaps: 10

US-09-985-689A-6 (1-434) x BQ622771 (1-640)

QY 19 LeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg--- 37
 Db 22 TTAAGAGGTAAAGTCAAGATATTGATGATGATGATGATGATGATGATGATGATGATGAT 81
 QY 38 -----AsnAspSer-----SerMetHisGluAlaPheArgGly 48
 Db 82 TTTCTTTTCAAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAGATCA 141

QY 49 LysileThrAlaileTyrrAlaLeuGlyArgThrAsnAsnAlaAspPro-----Asn 66
 Db 142 GTTGTAACTATATT-----ACTTACCATGACAATGAAGATTATGTAAT 186

QY 67 GlyHisGlyThrHisValalaglySerValLeuGlyAsnGly----- 80
 Db 187 GGTTCATGGTACACATGTTTGTGGCTCTGCAGAGGTACTCCAGAGGATCTTCATGGGCT 246

QY 81 ---ThrSerAsnlysglyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAsp 99
 Db 247 ATTTCATCATTTAGTGGTCTGTCACTGACGCAAGATGTCATTTATGAT---CTTCA 303

QY 100 SerAsnGlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrr 119
 Db 304 TCTGGAAGTTCTGAACCAACACCCGAGATACAGTCAATATGATCAAAACCATATAT 363

QY 120 SerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAla-----ProValAan 135
 Db 364 GATGCGAGGTGCAAGAGTACATGCTGATCTTGGGTTCTGTATCTTTCGCAAGTTATAT 423

QY 136 GlyAlaTyrrThrThrAspSerArgAsnValAspAspTyrrValarglys---AsnAspMet 154
 Db 424 GGTGGTATTCCGATGATGCTGGTGTATTTGATGCTATCTCTATGATGATCCAGAAATTC 483

QY 155 AlaValLeuPheAlaAlaGlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGly 174
 Db 484 TCTATCTAAGAGCTGCTGTAAT---AACGACTATTTCATCTTTATAGCTCAAGCA 540

QY 175 ThrAlaIysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 194
 Db 541 ACAGCTAAAAATGCAATTACAGTTGGTGTGTCGACAAACAGCTCATGTAATATGTTGTC 600

QY 195 -----TyrAlaAspAsnIleasn 200
 Db 601 GATGCAATGGAATATTGATTTCTCAGATAATGCTAAT 639

RESULT 3
 AK029048
 LOCUS
 DEFINITION
 Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched library, clone:473248M11 product:membrane-bound transcription factor protease, site 1, full insert sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 AK029048.1 GI:26325017
 HTG; CAP trapper.
 Mus musculus (house mouse)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 4
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 5
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 6
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 7
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 8
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 9
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 10
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 11
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 12
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 13
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 14
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 15
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 16
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 17
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 18
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 19
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 20
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 21
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 22
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 23
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 24
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 25
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 26
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 27
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 28
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 29
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 30
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 31
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 32
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 33
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 34
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 35
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 36
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 37
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 38
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 39
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 40
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 41
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 42
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 43
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 44
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 45
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 46
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 47
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 48
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 49
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 50
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 51
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 52
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 53
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 54
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 55
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 56
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 57
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 58
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 59
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 60
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 61
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 62
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 63
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 64
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 65
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 66
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 67
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 68
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 69
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 70
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 71
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 72
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 73
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 74
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 75
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 76
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 77
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 78
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 79
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 80
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 81
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 82
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 83
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 84
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 85
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 86
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 87
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 88
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 89
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 90
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 91
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 92
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 93
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 94
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 95
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 96
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 97
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 98
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 99
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 100
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 101
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 102
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 103
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 104
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 105
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 106
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 107
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 108
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 109
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 110
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 111
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 112
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 113
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 114
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 115
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 116
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 117
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 118
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 119
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 120
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 121
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 122
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 123
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 124
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 125
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 126
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 127
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 128
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 129
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 130
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 131
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 132
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 133
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 134
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 135
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 136
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 137
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 138
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 139
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 140
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 141
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 142
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 143
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 144
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 145
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 146
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 147
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 148
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 149
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 150
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 151
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 152
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 153
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 154
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 155
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 156
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 157
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 158
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 159
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 160
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 161
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 162
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 163
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 164
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 165
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 166
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 167
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 168
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 169
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 170
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 171
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 172
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 173
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 174
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 175
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 176
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 177
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 178
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 179
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 180
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 181
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 182
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 183
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 184
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 185
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 186
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 187
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 188
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 189
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 190
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 191
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 192
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 193
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 194
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 195
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 196
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 197
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 198
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 199
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 200
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 201
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 202
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 203
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 204
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 205
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 206
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 207
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 208
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 209
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 210
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 211
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 212
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 213
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 214
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 215
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 216
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 217
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 218
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 219
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 220
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 221
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 222
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 223
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 224
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 225
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 226
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 227
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 228
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 229
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 230
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 231
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 232
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 233
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 234
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 235
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 236
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 237
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 238
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 239
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 240
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 241
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 242
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 243
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 244
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 245
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 246
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 247
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 248
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 249
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 250
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 251
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 252
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 253
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 254
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 255
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 256
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 257
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 258
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 259
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 260
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 261
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 262
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 263
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 264
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 265
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 266
 AUTHORS
 TITLE

SVTKKAASWEGIAQGHIMITVAPAEELHSGAEHTSVKPIKVIITPPSRKVL
WQYHNLRYPPGYFPDNLKMDPLDWNHVDHVTNFRDMYOHLSMGYFVEVLGAPF
TCFDAQTGTLIVSEEEYFEEIAKLRDNDGLSVIFSDMTNTVMKRVKPYDE
NTROMPTGGANIPALNELLSVNMGFSGLYEGEFVLNHNMYASGCSIAKPE
KGVITQTFKDOGLEVLKQETAVENPILGYQIPSEGGRIVLGYDNCIDDSHRQ
KQFWLLDALLQVTSVGTVPBSLSHSGNEFORPPSGAGLAPPERMEGNHLRYSKVLEA
HGDPRPRLPACPHUSWAKPQELNETAPSNLWKQKLLSIDLKVLNPNRSNRPQV
RLESPGSGAWDIPGGIMPGRYNQEVGTIPVFALGNVALAFFVQISKAKSRPKR
RPRAPRQALQQAHPARTFSV
4180..4185
/note="putative"
polyA_signal
4198
/note="putative"
polyA_site
ORIGIN
Alignment Scores:
Pred. No.: 2e-11 Length: 4198
Score: 227.50 Matches: 115
Percent Similarity: 39.03% Conservative: 54
Best Local Similarity: 26.56% Mismatches: 177
Query Match: 10.11% Indels: 87
DB: 11 Gaps: 21
US-09-985-689A-6 (1-434) x AK029048 (1-4198)
QY 8 VallysAlaAspValAlaGlnSerSerTy:GlyLeuTyGlyGlnGlyGlnValAla 27
Db 1005 CTGAGGAGATGCTGTGG---CAGATGGATACACAGGTGCTAATGTCAGATTGCT 1061
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 1062 GTTTTGATATGGGCTC-----AGTGAGAAGCATCCGCATTTTAAG 1103
QY 48 GlyLysIleThrAlaIleTyAlaLeuGlyArgThrAsnAsnAlaAsn----- 63
Db 1104 AAT-----GTGAGGAGAGAACCACTGGACCATGAGCGGACCTG 1145
QY 64 ---AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThrSer 82
Db 1146 GATGATGGCTAGGCCATGGCACATTCGTTGCAGGT---GTGATTGCCAGCATGAGGGAG 1202
QY 83 AsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSerAsnGly 102
Db 1203 TGCCAAAGGATTGCTCCAGATGACAGCTGCACATCTTCAGGGTCTTTACCAACAT--- 1259
QY 103 GlyLeuGlyGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTySerAlaGly 122
Db 1260 -----CAGGTGCTTACACATCTTGGTTCTGGATGCTTCAACTATGCC 1304
QY 123 AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrThrAspSer 142
Db 1305 ATCCTAAAGAGATGGACGCTTCAACCTTAGCATCGTGCGCCGACCTTCATGATCAT 1364
QY 143 ArgAsnValAspAspTyx-----ValArgLysAsnAspMetAlaValLeuPheAlaAla 160
Db 1365 CGTTTGTGCAAGGTGTGGGAATTAACAGCTAACATGAATATGTTCTGCTATT 1424
QY 161 GlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1425 GGCAATGATGAGCTCTCTATGCGCACTCTGAATAACCTGCGATCAGATGATGATGATT 1484
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsn 200
Db 1485 GGNUGTGGTGGCATTCAC-----TTTGAAGATAACATC--- 1517
QY 201 HisValAlaGlnPheSerSerArgGly-----ProThrLysAspGly 214
Db 1518 -----GCTCGCTTTCTTCCAGGGGAATGACTACCTGGGAATTAACAGAGGCTATGGT 1571
QY 215 ArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeu 234
Db 1572 CGTGTGAAGCCTGACATGTGC-----ACCTATGGTCTGGAGTGGGGGTCCGGT 1622

QY 235 AlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyAlaTyMetGlyGlyThr 254
Db 1623 GTGAAAGGGGGCTGC-----CGTGCACTCTCAGGAC 1655
QY 255 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLys 274
Db 1656 AGTGTGCTTCCCAGTGTGCTGGCGCTCACCTTGTAGTAAACACAGTACACAG 1715
QY 275 AsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaThr 294
Db 1716 CGGAGAGCTGGTG-----AATCTGCCAGTGTGAGCAAGCTTTGATAGCTCAGCCGG 1769
QY 295 AspIleGlyLeuGlyTyProSerGlyAsnGlnGlyTrpGlyArgValThrLeuAspLys 314
Db 1770 AGACTTCTCT---GGGTCAACATGCTGAGCAAGTTCATGGCAAGTGTGCTCGCA 1826
QY 315 SerLeuAsnValAlaPheValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThr 334
Db 1827 GCTTATCAGATC-----CTCAGCAGCTATAAACCGCAGGACGCTGAGT 1871
QY 335 TyrSerPheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValTrp-----Ser 352
Db 1872 CCTAGCTACATGACCTGACTGAGTGTCCC-----TACATGTGCGCTTACTGTC 1919
QY 353 AspAlaProAlaSerThrSerAlaSerValThrLeuValAsnAspLeuAspLeuValIle 372
Db 1920 TCCAGCCTTACTTACTATGAGGAATGCCAACATCGTTAAT-----GTC 1964
QY 373 ThrAlaProAsnGlyThrLysTyValGlyAsnAspPheThrAlaProTyAspAsnAsn 392
Db 1965 ACCATCTCAATGGCATGGGTGTCACAGGAAGATTGTGATAAGCCT-----GAG 2015
QY 393 Trp-----AspGlyArgAsnAsnValGluAsnValPhe----- 403
Db 2016 TGGCGACCTATTTTACCAGATGGAGACATTTGAAGTGGCTTCTCTACTCTCTCA 2075
QY 404 IleAsnAlaProGlnSerGlyThrTyThrValGluVal 416
Db 2076 GTGTTGTGGCTGTGCTAGTTTACCTGCCATCTCCATT 2114

RESULT 4
TA319G10P 532 bp DNA linear GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 319g10, forward sequence,
DEFINITION genomic survey sequence.
ACCESSION AL492464.1 GI:11867408
VERSION AL492464.1
KEYWORDS GSS.
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei
REFERENCE 1 (bases 1 to 532)
AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
Melville,S.B., Rajandream,M.A. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, B-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 Gutat 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nalsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES

Location/Qualifiers
 1. 532
 /organism="Trypanosoma brucei"
 /mol_type="genomic DNA"
 /strain="TREU92"
 /db_xref="taxon:5691"
 /clone="319g10"

ORIGIN

Alignment Scores:
 Pred. No.: 1.57e-11 Length: 532
 Score: 213.50 Matches: 61
 Percent Similarity: 48.37% Conservative: 28
 Best Local Similarity: 33.15% Mismatches: 68
 Query Match: 9.49% Indels: 27
 DB: 29 Gaps: 6

US-09-985-689A-6 (1-434) x TA319G10P (1-532)

QY 93 ValPheGlnSerValMetAspSerAsnGlyLeuGly----- 105
 Db 2 GTAATCCCAAGTATATGTTGCGCCAGGGGCGAAGATTCTTCAGGGGTTGGGCTGCC 61
 QY 106 -----GlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyr 119
 Db 62 CATCCAGTCAGAGCTTGCTCCCTCCAGCAGCTTACTCAAAATATTCCTCCCGGTATAT 121
 QY 120 SerAlaGlyAlaArgIleHisThrAsnSerTTPGlyAlaProValAsnGlyAlaTyrThr 139
 Db 122 GCGCTGGAGCCGCTGTTCTCAAACTCGTGGGTTTTGTTGCTCCCTCCAGTATTC 181
 QY 140 ThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeu---Phe 158
 Db 182 GCTGTGAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241
 QY 159 AlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsn 178
 Db 242 TCACCTGGCAACAGTATCCAGATGGC-----CTAATGACTCCGTCGTCGTAAGAAC 295
 QY 179 AlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsn 198
 Db 296 GTGATGTGCTGGGTGCACAAAACCGTG-----TTTGACGCTTCGAAAGAC----- 343
 QY 199 IleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysPro 218
 Db 344 -----ATTGTTCTTCGTTTCTTCATGTCCTTCCCTCTGCTGCTGCTGCTGCTGCTGCT 397
 QY 219 AspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer 238
 Db 398 GATCTTGTGCTGCCGGGGAAGAGTGGTGTCTGCTCTTCTCTCTGCTGCTGCTGCTGCTGCT 457
 QY 239 SerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThr 258
 Db 458 AAA-----CAATGTAAAGTGGTGGCCAGCGGNGTTCATCGATGGCACT 502
 QY 259 ProIleValAla 262
 Db 503 GCGGCCGCTCGG 514

RESULT 5
 BJ393752
 LOCUS
 DEFINITION BJ393752 Dictyostelium discoideum cDNA library, SF Dictyostelium
 discoideum cDNA clone dds32b16 5', mRNA sequence.

ACCESSION BJ393752
 VERSION 1
 KEYWORDS EST.

SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

REFERENCE 1 (bases 1 to 594)

AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.

TITLE Full length cDNA of Dictyostelium discoideum at the slug stage

JOURNAL COMMENT

Unpublished (2002)
 Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.

FEATURES

Location/Qualifiers
 1. 594
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AX4"
 /db_xref="taxon:44889"
 /clone="dds32b16"
 /sex="mat A"
 /dev stage="slug stage"
 /clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN

Alignment Scores:
 Pred. No.: 1.9e-11 Length: 594
 Score: 213.50 Matches: 55
 Percent Similarity: 50.81% Conservative: 39
 Best Local Similarity: 29.73% Mismatches: 62
 Query Match: 9.49% Indels: 29
 DB: 12 Gaps: 8

US-09-985-689A-6 (1-434) x BJ393752 (1-594)

QY 49 LysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspProAsnGlyHis 68
 Db 21 AAAGTTGTAATTATATACCATCAACAAAGCAGCATAGTGTATAAAGTGGATGTCAC 80
 QY 69 GlyThrHisValAlaGlySerValLeuGly-----AsnGlyThr 81
 Db 81 GGTACACATATTTGTGTTCTTCGACAGGAGTCTCCAGAGGATTCCTCAGTAAATATTTCA 140
 QY 82 SerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSerAsn 101
 Db 141 TCATTTAGTGTCTGCACTGATGCAAGATTGCATCTTCTGTTTGGCAAGTGGTTCA 200
 QY 102 GlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSerAla 121
 Db 201 TCAAGTTTGACACCT---CCATCGGATTGAAACATATATATCAACCATTTATATGACGCA 257
 QY 122 GlyAlaArgIleHisThrAsnSerTrpGlyAla-----ProValAsnGlyAla 137
 Db 258 GGTGCAAGAGTGCATTGTGATTTCTGGGTTCTGTATCAGTAGAGGGGTATACAGGTAGT 317
 QY 138 TyrThrAspSerArgAsnValAspTyrValArgLysAsn---AspMetAlaVal 156
 Db 318 TATTCATCAGACACTGCTTCAATTCATGATTCCTTTTCACTCATCCAGATTTTCATCATT 377
 QY 157 LeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSer-----AlaProGly 174
 Db 378 CTTAGAGCTGTGGTAAC-----AACGAGCAATACCTATCTACTACTCACTCAATCC 428
 QY 175 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg----- 189
 Db 429 ACTGCAAGAGATGTTATTACCGTTGGTCTCATCAACAATTCATGAAATATTATTA 488
 QY 190 -----ProSerPheGlySerTyrAlaAspAsnIleAsn-----HisValAla 203
 Db 489 GATGTCCTCAAAATATATAATATATATCAATCATCTCTCGATATAAATCAAGAGTTAATGT 548
 QY 204 GlnPheSerSerArg 208
 Db 549 GATTTCGATAGCAGA 563

RESULT 6

BJ142519/c

LOCUS

2141 bp mRNA linear EST 24-APR-2002

3

KEYWORDS	EST.	Mus musculus (house mouse)	1. .771	Location/Qualifiers	1. .771	Source
SOURCE	Mus musculus			/organism="Mus musculus"		
ORGANISM	Mus musculus			/mol_type="mRNA"		
REFERENCE	1	(bases 1 to 771)		/strain="CS7BL/6"		
AUTHORS	NIH-MGC	http://mgc.nci.nih.gov/		/db_xref="taxon:10090"		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			/clone="IMAGE:6816072"		
JOURNAL	Unpublished (1999)			/issue_type="whole brain"		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: c9apbs-remail.nih.gov Tissue Procurement: Dr. Jim Lin, University of Iowa cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov This clone was contributed by the Brain Molecular Anatomy Project (BMAP)			/lab_host="PH10B (T1 phage resistant)"		
FEATURES	Seq primer: pyX-5.			/clone_lib="NIH BMAP FWO"		
source	1. .771			/notes="Organ: Brain; Vector: pyX-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into pyX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."		
ORIGIN	US-09-985-689A-6 (1-434) x CA320325 (1-771)					
Alignment Scores:						
Pred. No.:	7.42e-10	Length:	771			
Score:	200.00	Matches:	83			
Percent Similarity:	41.26%	Conservative:	35			
Best Local Similarity:	29.02%	Mismatches:	112			
Query Match:	8.99%	Indels:	56			
DB:	14	Gaps:	12			
US-09-985-689A-6 (1-434) x CA320325 (1-771)						
Qy	8	VallyAspValAlaGlnSerSerTyGlyLeuTyGlyGlnGlyGlnValValAla	27			
Db	40	CTGAGGAGATGTCGTGGG---CAGATGGGATACACAGGTGCTCTAATGTCAGAGTGTCT	96			
Qy	28	ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg	47			
Db	97	GTTTTGTACTGGGCTC-----AGTGAGAGCATCCGCATTTTAG	138			
Qy	48	GlyLysIleThrAlaIleTyThrAlaLeuGlyArgThrAsnAsnAla	63			
Db	139	AAT-----GTGAAGAGAGAGAACCACTCGACCAATGAGCGGACCCCTG	180			
Qy	64	---AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThrSer	82			